

# Reliability of the Berg Balance Scale as a Clinical Measure of Balance in Community-Dwelling Older Adults with Mild to Moderate Alzheimer Disease: A Pilot Study

Susan W. Muir-Hunter, PT, PhD;<sup>\*†</sup> Laura Graham, PT, PhD;<sup>‡</sup>  
Manuel Montero Odasso, MD, PhD<sup>†§¶</sup>

## ABSTRACT

**Purpose:** To measure test–retest and interrater reliability of the Berg Balance Scale (BBS) in community-dwelling adults with mild to moderate Alzheimer disease (AD). **Method:** A sample of 15 adults (mean age 80.20 [SD 5.03] years) with AD performed three balance tests: the BBS, timed up-and-go test (TUG), and Functional Reach Test (FRT). Both relative reliability, using the intra-class correlation coefficient (ICC), and absolute reliability, using standard error of measurement (SEM) and minimal detectable change (MDC<sub>95</sub>) values, were calculated; Bland–Altman plots were constructed to evaluate inter-tester agreement. The test–retest interval was 1 week. **Results:** For the BBS, relative reliability values were 0.95 (95% CI, 0.85–0.98) for test–retest reliability and 0.72 (95% CI, 0.31–0.91) for interrater reliability; SEM was 6.01 points and MDC<sub>95</sub> was 16.66 points; and interrater agreement was 16.62 points. The BBS performed better in test–retest reliability than the TUG and FRT, tests with established reliability in AD. Between 33% and 50% of participants required cueing beyond standardized instructions because they were unable to remember test instructions. **Conclusions:** The BBS achieved relative reliability values that support its clinical utility, but MDC<sub>95</sub> and agreement values indicate the scale has performance limitations in AD. Further research to optimize balance assessment for people with AD is required.

**Key Words:** accidental falls; Alzheimer disease; dementia; postural balance; reproducibility of results.

## RÉSUMÉ

**Objectif :** Mesurer la fiabilité test-retest et interévaluateurs de l'échelle de Berg chez les adultes âgés qui résident dans la collectivité et qui sont atteints de la maladie d'Alzheimer au stade léger à modéré. **Méthode :** Un échantillon de 15 adultes (âge moyen de 80,20 [ET de 5,03] ans) atteints d'Alzheimer ont effectué trois tests d'équilibre : l'échelle de Berg (*Berg Balance Scale*), le test de lever et marcher chronométré (*timed up-and-go* [TUG]) et le test de portée fonctionnelle (*Functional Reach Test*). La fiabilité relative, à l'aide du coefficient de corrélation intraclass (CCI) et la fiabilité absolue, à l'aide de l'erreur type de mesure (ETM) et des valeurs de changement détectable minimal (CDM<sub>95</sub>), ont été calculées; des courbes de Bland et Altman ont été construites pour évaluer la convergence interévaluateurs. L'intervalle de test-retest était d'une semaine. **Résultats :** En ce qui concerne l'échelle de Berg, les valeurs de fiabilité relative étaient de 0,95 (95% CI, 0,85 à 0,98) pour la fiabilité test-retest et de 0,72 (95% CI, 0,31 à 0,91) pour la fiabilité interévaluateurs; l'ETM était de 6,01 points et le CDM<sub>95</sub> était de 16,66 points, tandis que la convergence interévaluateur était de 16,62 points. L'échelle de Berg a obtenu de meilleurs résultats que les tests de lever et marcher chronométré et de portée fonctionnelle, tests ayant une fiabilité établie pour l'Alzheimer en fiabilité test-retest. De 33% à 50% des participants avaient besoin de plus de directives que les instructions normalisées, car ils n'étaient pas capables de se souvenir des instructions de test. **Conclusions :** L'échelle de Berg a obtenu des valeurs de fiabilité relative qui appuient son utilité clinique, mais les valeurs de CDM<sub>95</sub> et de convergence indiquent que cette échelle présente des limites en matière de rendement pour les cas d'Alzheimer. D'autres recherches visant l'optimisation de l'évaluation de l'équilibre chez les personnes atteintes d'Alzheimer doivent être effectuées.

From the: \*School of Physical Therapy; §Schulich School of Medicine & Dentistry, Division of Geriatric Medicine, and ¶Department of Epidemiology and Biostatistics, University of Western Ontario; †Gait and Brain Lab, Parkwood Hospital, Lawson Health Research Institute, and ‡Neurotrauma Rehab Program, Parkwood Hospital, London, Ont.

**Correspondence to:** Susan W. Hunter, School of Physical Therapy, University of Western Ontario, Rm. 1588, Elborn College, London, ON N6G 1H1; susan.hunter@uwo.ca.

**Contributors:** All authors designed the study; collected, analyzed, and interpreted the data; drafted or critically revised the article; and approved the final draft.

**Competing Interests:** None declared.

This research was supported by the WCA Committee through Alzheimer Outreach Services of McCormick Home. Dr. Muir-Hunter was supported by an Alzheimer Society of Canada Post-Doctoral Award and Ontario Research Coalition Early Researcher Award. The Gait and Brain Lab at the University of Western Ontario is supported by funding from the Canadian Institutes for Health Research (CIHR), the Ontario Ministry of Research and Innovation, and the Department of Medicine Program of Experimental Medicine (POEM).

**Acknowledgements:** The authors thank Magdalen Carter, Director of Alzheimer Outreach Services of McCormick Home; Steve Crawford, CEO, WCA; and the staff and clients at the Alzheimer Outreach Services day program for their hospitality, assistance in organizing this project, and participation in the data collection process.

*Physiotherapy Canada* 2015; 67(3);255–262; doi:10.3138/ptc.2014-32

Reliability of outcome measures is important for clinical practice because the consistency or repeatability of measures affects how recorded scores are interpreted to determine whether a true change in function has occurred. Outcome measures developed and evaluated for reliability in one patient population should be reviewed in other patient groups that present with unique features that could potentially alter a test's performance.

Dementia is a set of diseases, each with a distinct underlying pathology; although these diseases have commonalities, clinical features also distinguish one subtype from another. More important, gait and balance problems have been found to vary considerably by dementia subtype, and therefore it is important to be aware of the heterogeneity in functional ability.<sup>1</sup> Some of the existing literature on the reliability of physical performance measures in people with dementia is limited by the use of a sample consisting of participants with a non-specific diagnosis of dementia and no quantification of disease severity or disease sub-type.<sup>2-6</sup>

Falls are a major cause of disability and dependence in older adults, particularly in those with dementia:<sup>7</sup> People with dementia have an annual fall risk of 60%–80%,<sup>8</sup> twice that of community-dwelling older adults with normal cognition. People with dementia who fall are also 5 times more likely to be admitted to an institution than those who do not.<sup>9</sup> Finally, the cognitive deficits associated with dementia affect the ability to perform activities of daily living independently.<sup>10</sup> The number of people living with dementia is expected to increase dramatically in the coming years; in addition to indirect costs and caregiver burden, the direct costs of emergency, acute care, rehabilitation, and long-term care in this population are projected to be substantial for the health care system.<sup>10</sup>

Alzheimer disease (AD), the most common type of dementia, is characterized by hallmark changes in memory and language that represent potential barriers to the assessment of physical function. These changes, which begin in the earliest phases of the disease, create difficulty in understanding spoken words and following multistep instructions; they progress over time with increasing disease severity to adversely affect the performance of physical tasks.<sup>11</sup> Taken together, these changes could compromise the ability of outcome measures to accurately quantify or detect change over time in a physiotherapy treatment program.<sup>12-14</sup> Therefore, when outcome measures are used with this population, the content of test instructions needs to be aligned with known communication strategies that are responsive to language changes in AD, and the number of steps required to execute a given task should be minimized.

When assessing people with a diagnosis of AD, it is important to appreciate whether the clinical test measures balance function (i.e., both static and dynamic

abilities) or whether it instead measures how well a person with compromised cognitive function (specifically, limitations in short-term memory and executive function) can follow multi-step instructions and understand complex language in the instructions. An inability to follow multi-step commands may not change with a rehabilitation intervention; conversely, if balance function does improve, the improvement will not necessarily be identified if cognitive impairment prevents quantification of abilities on the standardized test. Among the available physical performance measures regularly used in clinical practice, only a small selection have demonstrated reliability (test-retest reliability) in people with AD (the timed up-and-go [TUG] test,<sup>2,3,12,14</sup> Functional Reach Test [FRT],<sup>12</sup> 6-minute walk test,<sup>13,14</sup> and Physical Performance Test<sup>15</sup>); investigation of additional outcome measures has the potential to increase the assessment options for therapists.

The Berg Balance Scale (BBS), developed as a clinical measure of functional balance specifically for use with older adults, is one of the scales most commonly used in clinical practice by physiotherapists.<sup>16</sup> Suggested applications include comparing balance between groups of people, describing balance in an individual, monitoring balance ability over time, and evaluating the effectiveness of rehabilitation treatment.<sup>17</sup> Among Ontario physical therapists practising in geriatrics, the BBS and the TUG are the two most commonly used measures of functional performance.<sup>16</sup> Because the BBS evaluates 14 tasks, it has the potential to provide a greater overview of abilities than the TUG; also, the graduated nature of the tasks provides safety in the assessment process and can assist in treatment planning. The psychometric properties of the BBS have been well demonstrated in many patient populations (people with stroke, multiple sclerosis, acquired brain injury, Parkinson disease, and spinal cord injury), but it has not been evaluated among people with AD.

Older adults with dementia can present unique challenges in the rehabilitation setting,<sup>18,19</sup> but they are able to participate and make gains in mobility through structured rehabilitation interventions.<sup>20</sup> Cognitive impairment is a particularly important factor for physiotherapists to understand, because its prevalence among patients in geriatric inpatient rehabilitation services has been reported to range from 31% to 45%.<sup>21-23</sup> It is very important that scales commonly used by physiotherapists to evaluate and assess balance function have demonstrated acceptable reliability. Our objective in this study, therefore, was to measure the test-retest and interrater reliability of the BBS in a population of community-dwelling older adults with mild to moderate AD. We hypothesized that the BBS would be reliable and would perform well relative to other balance scales with established reliability, specifically the TUG and the FRT, in this population.

## METHODS

### Participants

We recruited a convenience sample of adults with a diagnosis of mild to moderate AD from a day program for community-dwelling older adults with dementia. Referral to the program is based on a confirmed diagnosis of dementia by a geriatrician according to the criteria of the National Institute of Neurologic and Communicative Disorders and Stroke—AD and Related Disorders Association.<sup>24</sup> Participants were included if they were age 65 years or older, medically stable, English speaking, and able to understand simple instructions. Potential participants were excluded if they had any neurological, musculoskeletal, or cardiorespiratory impairment that could compromise safe administration of the testing protocol. The study was approved by the University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects. All participants or their caregivers provided written informed consent before participation in the study.

### Measures of balance

Three clinical tests of balance were evaluated: the BBS, TUG, and FRT. Participants wore their usual footwear for all balance tests.

### Berg Balance Scale

The BBS consists of 14 functional tasks of increasing difficulty, each scored on a scale ranging from 0 to 4 (0 = unable to perform the task; 4 = task is performed independently).<sup>17,25–27</sup> The maximum possible score is 56, indicating no identifiable balance difficulties.

### Timed up-and-go test

The TUG assesses mobility and locomotor performance in older adults.<sup>28,29</sup> In this study, participants were timed from the moment they were instructed to stand up from a chair (seat height = 48 cm), walk at their preferred usual pace for 3 m, turn around, and walk back to sit in the chair. Participants were allowed to use the arms of the chair to stand up and to use their usual mobility aid, if any. Bischoff and colleagues<sup>30</sup> have proposed that women ages 65–85 years should be able to complete the TUG in 12 seconds or less. Test–retest reliability has been established for this scale in older adults with mild to moderate AD.<sup>12,14</sup> Instructions for performance of the test were given according to the information provided by Ries and colleagues<sup>14</sup> in their study of older adults with AD.

### Functional Reach Test

The FRT was developed to provide a functional quantification of the limits of stability,<sup>31</sup> defined as the maximal distance one can lean forward, backward, and sideways while standing without changing one's foot position.<sup>31</sup> The distance that people can lean decreases with age, which is believed to reflect compensation for

impaired postural control mechanisms.<sup>31</sup> A distance of less than 20.3 cm has been associated with an increased fall risk in community-dwelling older adults.<sup>32</sup> Test–retest reliability has been established for this scale in older adults with mild to moderate AD.<sup>12</sup>

### Testing procedure

To evaluate the reliability of the functional outcome measures, participants attended two testing sessions scheduled 1 week apart. This time frame was chosen because balance variables could reasonably be expected not to change during this time, in the absence of a fall or an acute illness. The order of the balance tests and the raters was randomized. Participants were allowed a short rest between evaluations, as needed. Two physiotherapists (SWH, LG) with experience in assessing and treating older adults with balance problems performed all the evaluations.

In the first session, assessments were carried out by both therapists (SWH, LG); in the second session, all assessments were completed by one therapist (SWH). Instructions for the performance of each task in the BBS used the standardized wording that accompanies the tool; the assessing therapists underwent a training session to ensure consistency in the administration of all tests. For the purposes of the study, *cueing* was defined as providing any additional verbal, visual, or tactile direction necessary to ensure correct performance of the task after the initial set of standardized instructions was given. Cueing was considered different from *supervision*, defined as the use of standby physical assistance to ensure safe performance of a task in case the participant lost his or her balance. Assessors were instructed to record either *yes* or *no* on the test form regarding whether cueing was required beyond the standardized verbal instructions; for the BBS, each individual task had cueing notes. The need for cueing was evaluated during test session 1 (the interrater reliability session), the first time the test was administered.

### Additional outcome measures

Global cognitive status was assessed using the Mini-Mental State Examination (MMSE), which assesses orientation, attention, memory, and language and has been validated as a tool with high reliability in a variety of patient populations.<sup>33</sup> Severity of dementia was categorized by MMSE score as mild (MMSE > 20 points), moderate (MMSE = 10–20 points), or severe (MMSE < 10 points).<sup>34</sup>

### Data analysis

Relative reliability is the degree to which individuals maintain their position in a sample when repeated measurements are performed.<sup>35</sup> We calculated two measures of relative reliability for each balance test: interrater reliability (the degree to which the measurement tool yields similar results at the same time with more

than one assessor) and test–retest reliability (the degree to which a result on one instrument is equivalent to the result on the same instrument across days). Relative reliability was quantified using the intra-class correlation coefficient (ICC). An ICC value of more than 0.9 was considered excellent, values between 0.8 and 0.9 were considered good, values between 0.7 and 0.8 were considered fair, and values less than 0.7 were considered to be of questionable clinical value.<sup>36</sup>

Absolute reliability is the degree to which repeated measurements of the same tool vary for an individual; smaller variation indicates higher reliability.<sup>35</sup> Two measures of absolute reliability were quantified from the two evaluations done by the same rater: the standard error of measurement (SEM), an expression of measurement error in the same units as the scale, calculated as  $SEM = SD\sqrt{(1-ICC)}$ ,<sup>37</sup> and the minimum detectable change ( $MDC_{95}$ ), an estimate of the smallest change in the score that can be detected beyond measurement error, calculated as  $MDC_{95} = SEM \times \sqrt{2} \times 1.96$ .<sup>37</sup> In addition, we constructed Bland–Altman plots to measure agreement between the two assessors for each balance test.<sup>38</sup> A Bland–Altman plot involves graphing the difference in balance scores between the two testing sessions against the mean of the sample balance scores. The data on the use of additional cueing were quantified with descriptive statistics only. Using the method described by Walter and colleagues,<sup>39</sup> we calculated a required sample size of 12 to identify a desired ICC of 0.9, with a lower CI of 0.60, given  $\alpha = 0.05$  and  $\beta = 0.20$ . All statistical analyses were performed using IBM SPSS Statistics version 21.0 (IBM Corporation, Armonk, NY).

## RESULTS

A convenience sample of 15 older adults consented to participate in the study. The sample consisted of 11 men and 4 women; the mean age was 80.20 (SD 5.03) years. The mean MMSE score was 20.0 (SD 5.5), which indicates mild to moderate disease severity; participants had a mean 11.7 (SD 3.4) years of education, and none used a mobility aid.

All participants were able to complete all of the balance tests; no tests were terminated because of safety concerns, on participant request, or because the participant, even with additional cueing, could not understand and execute each task. However, some participants required additional cueing for instructions on some tests; 33% required cueing for the TUG because they could not remember all the steps in the test's instructions, and cueing was required for the following BBS items for the same reason: item 5, transfers (33%); item 8, limits of stability–reaching forward (25%); item 10, turning to look behind left and right shoulder in standing (50%); item 11, turn 360° (50%); item 12, placing alternate foot on step while standing unsupported (33%); and item 13,

**Table 1** ICCs for Test–Retest and Interrater Reliability Values for Community-Dwelling Older Adults with Mild to Moderate Alzheimer Disease ( $n = 15$ )

Balance test	ICC	95% CI	$p$ -value*
Test–retest reliability			
BBS	0.95	0.85–0.98	<0.0001
TUG	0.72	0.33–0.90	0.002
FRT	0.81	0.52–0.94	<0.0001
Interrater reliability			
BBS	0.72	0.31–0.91	0.002
TUG	0.98	0.93–0.99	<0.0001
FRT	0.79	0.43–0.94	0.001

\*Statistical significance was set at  $p < 0.05$ .

ICC = intra-class correlation coefficient; BBS = Berg Balance Scale;

TUG = timed up-and-go test; FRT = Functional Reach Test.

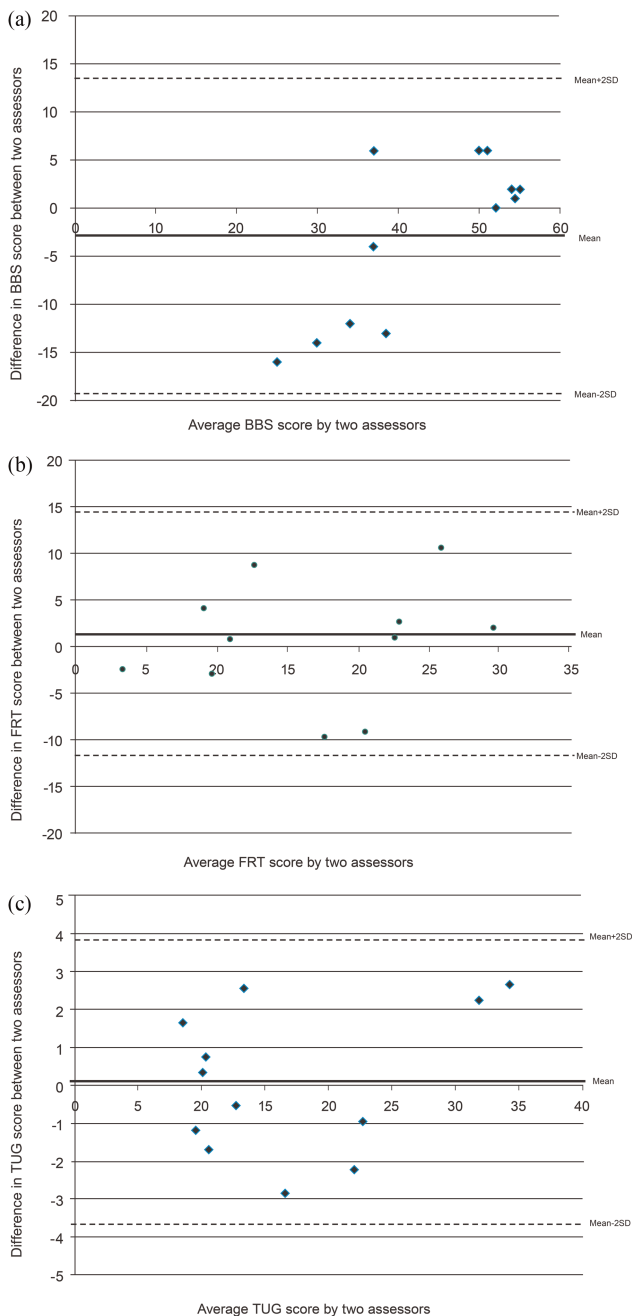
tandem standing unsupported with one leg in front (17%).

The BBS showed excellent test–retest reliability (ICC = 0.95; 95% CI, 0.85–0.98;  $p < 0.0001$ ), performing better than both the TUG and the FRT; all three tests achieved test–retest values indicating clinical utility (ICC  $\geq 0.70$ ). The interrater reliability of the BBS was lower than its test–retest reliability and lower than those of the TUG and FRT; nevertheless, all three tests achieved interrater reliability values that indicated clinical utility (ICC  $\geq 0.70$ ; see Table 1).

The results of relative and absolute test–retest reliability analyses are presented in Table 2. For the BBS, the SEM was 6.01 points; more important, however, the  $MDC_{95}$  was 16.66. The Bland–Altman plot for the BBS (Figure 1) also shows sufficient disagreement between the two evaluators in the same session to be clinically important ( $\pm 16.62$  points) and to cause problems with interpretation. The mean BBS score improved by an average of 3 points between the trials for interrater reliability (see Table 2); no such improvement occurred with the TUG or the FRT.

## DISCUSSION

Our findings show that although the BBS demonstrates acceptable relative reliability in both test–retest and interrater evaluation, measures of absolute reliability and agreement suggest that the scale has limitations when used with adults with mild to moderate AD. These limitations mean that differences in scores obtained by multiple therapists evaluating the same person would be large enough to make interpretation of balance status difficult, if not impossible, limiting the therapist's ability to determine whether the client's ability has truly changed over time. However, the BBS showed better relative reliability than the TUG and the FRT, both measures with established reliability in mild to moderate AD.<sup>12,14</sup> Measures of reliability for the TUG were consistent with results reported by Ries and colleagues,<sup>14</sup> based on a



**Figure 1** Bland–Altman plots for agreement between two physiotherapists performing the (a) Berg Balance Scale (BBS), (b) Functional Reach Test (FRT), and (c) timed up-and-go test (TUG) on community-dwelling older adults with mild to moderate Alzheimer disease.

sample of people with mild to severe AD; Suttanon and colleagues<sup>12</sup> reported lower values for people with mild to moderate AD. No previous study of reliability in people with AD has published measures of agreement; our study thus adds valuable information on the variation that can be present between assessments using the TUG, FRT, and BBS.

To the best of our knowledge, ours is the first study to evaluate the BBS in a group of older adults with a diagnosis of dementia, including a specific quantification of both types of disease and disease severity (i.e., mild to moderate AD). Previous research has demonstrated that in a sample of older adults living in a personal care home setting (i.e., in their own apartments with care staff available 24 hours a day for assistance with activities of daily living), the BBS has acceptable test–retest reliability (ICC = 0.77) over a retest period of 1–2 weeks.<sup>40</sup> Unfortunately, this research did not present information on participants' cognitive status, either via global cognitive testing scores or a dementia diagnosis, so the results cannot be directly compared with ours.

It is also important to note that the other balance scales evaluated in our study, the TUG and the FRT, achieved acceptable interrater reliability. Previous research on the psychometric properties of these balance scales has been conducted in a mixed population of patients with mild to moderate AD or a sample with mixed dementia types.<sup>5,12</sup> The relative proportions of people with mild and moderate disease could affect overall test performance between samples, and therefore this information should be included in study sample descriptions because test performance may decrease with increasing disease severity. Previous evidence on the FRT's reliability has been contradictory and limits knowledge translation recommendations.<sup>12,41</sup> Because our reliability results for the FRT are at the threshold value of clinical utility, it is recommended that it not be used as the sole measure of balance assessment for adults with mild to moderate AD. Our study sample was distinct from those of other studies<sup>12,14</sup> in that although all participants lived in the community, they were also enrolled in an outpatient day program, which suggests their disease had progressed to the point at which their caregivers were seeking respite. No information was available on the presence of responsive behaviours or the reasons for families' initiating day program attendance.

None of the previous research on using balance measures with this population has reported on the percentage of people who did or did not require cueing to perform the test or the individual items within balance scales such as the BBS. Our study is thus a novel contribution to the literature and highlights some of the limitations that people with AD can experience with multi-step commands.

The use of cueing may not necessarily be detrimental in the evaluation of complex tasks, but tasks scored on the basis of completion time are unfairly biased against people with memory and language problems. Another important consideration is that providing cues during a task has the potential to adversely affect performance, such that this type of compensation by the assessor may still not allow for an accurate quantification of ability. A person with AD who needs to listen to directions from

**Table 2** Scores, Test–Retest Reliability, Standard Error of Measurement, and Minimal Detectable Change of Four Balance Measures Evaluated on Two Occasions in People with Mild to Moderate Alzheimer Disease ( $n = 15$ )

	BBS	TUG, s	FRT, cm
Trial 1, mean (SD)	41.67 (14.25)	16.94 (9.22)	19.00 (11.22)
Trial 2, mean (SD)	44.67 (7.84)	16.88 (8.65)	17.68 (8.97)
No. (%) who required cueing to complete tasks during testing	8 (53.3)	5 (33.3)	3 (20.0)
ICC (95% CI)	0.95 (0.85–0.98)	0.72 (0.33–0.90)	0.81 (0.52–0.94)
SEM	6.01	1.24	4.56
MDC <sub>95</sub>	16.66	3.44	12.64
Agreement (Bland–Altman plot)	16.94	3.82	16.94

BBS = Berg Balance Scale; TUG = timed up-and-go test; FRT = Functional Reach Test; ICC = intra-class correlation coefficient; SEM = standard error of measurement; MDC<sub>95</sub> = minimal detectable change.

the evaluator is performing a superimposed attention-demanding task that may diminish task performance, creating an inadvertent dual-task test scenario; a timed test will take longer if the person must wait to be told the next component in the sequence of movements. In addition, the complexity of the testing could vary from one testing occasion to another, or between individuals, if the amount of cueing is not consistent. Therefore, it is important that clinicians note both the level and the form of cueing (i.e., verbal, tactile, or visual) required during testing, to allow replication of testing conditions and so that any changes in the level of guidance required can be documented.

Finally, scoring for the BBS task “Turn 360°” requires that points be deducted if cueing is required, which raises issues of validity because a person who requires cueing for this task may not necessarily show poorer balance during daily activities, when cues are internal and environmental. The wording of standardized instructions and the scoring options may need to be reviewed to better quantify ability by optimizing communication strategies for people with AD.

Our study has several important limitations. First, the sample size was small, and a lack of statistical power may have had an adverse effect on results; therefore, the study should be repeated with a larger sample. Second, the 1-week separation between testing could be viewed as a limitation because this time frame may be seen as not stable and is therefore a source of discrepancy. In comparison with other studies, the methodology for test–retest evaluation used in this study was comparable with respect to time frame for retesting, and it is not a factor for differences between studies. Third, the study did not have an equal distribution of people with mild and moderate AD, preventing a stratification of results by disease severity that would provide greater understanding of performance on the BBS.

A strength of our study was the recording of additional cueing and the recognition of its potential detrimental impact (by creating a dual-task challenge) as

well as the fact that language and memory impairments can compromise test performance. In addition, our study sample consisted entirely of people with a confirmed diagnosis of mild to moderate AD, which limits the variation among individuals that a heterogeneous sample of multiple dementia sub-types can present. As a pilot study, it identified avenues to explore in future research and the need to provide greater quantification of cueing or participant difficulties with tasks to enhance performance on the BBS by people with AD.

## CONCLUSIONS

This study demonstrated that the BBS possesses clinically useful values of relative reliability (i.e., test–retest and interrater reliability), but that measures of absolute reliability and agreement show limitations in test performance in people with mild to moderate AD. The complex language and multi-step nature of the standardized instructions for the BBS should be reviewed for possible modifications to help overcome barriers to the assessment of balance ability. Our study provides further evidence that the TUG is an acceptable test for people with AD, but it is still important to note the need for additional cueing to complete the components of the test activity and to be aware that this may adversely affect performance rather than compensating for cognitive deficits challenged by the test. More work is required to provide clinicians with an acceptable complement of testing options for people with AD, across the spectrum of disease severity, to quantify balance and measure change with interventions.

## KEY MESSAGES

### What is already known on this topic

Balance function is compromised in people with AD, placing them at an increased risk for falls and serious fall-related injuries. The reliability of most commonly performed balance assessment scales has not been established for this population, which may limit therapists' ability to accurately quantify and monitor change over

time with a rehabilitation intervention. The BBS, one of the scales most commonly used by physiotherapists, has not been evaluated in this population.

### What this study adds

This study demonstrated that the BBS achieved relative reliability values that support clinical utility, but the MDC<sub>95</sub> and agreement values indicate that the scale has performance limitations in mild to moderate AD. The BBS should be reviewed to address potential barriers in the administration of the test, such as language and short-term memory limitations and difficulties in following multistep commands that are hallmarks of AD.

### REFERENCES

- Allan LM, Ballard CG, Burn DJ, et al. Prevalence and severity of gait disorders in Alzheimer's and non-Alzheimer's dementias. *J Am Geriatr Soc.* 2005;53(10):1681–7. <http://dx.doi.org/10.1111/j.1532-5415.2005.53552.x>. Medline:16181166
- Blankevoort CG, van Heuvelen MJG, Scherder EJA. Reliability of six physical performance tests in older people with dementia. *Phys Ther.* 2013;93(1):69–78. <http://dx.doi.org/10.2522/ptj.20110164>. Medline:22976448
- Thomas VS, Hageman PA. A preliminary study on the reliability of physical performance measures in older day-care center clients with dementia. *Int Psychogeriatr.* 2002;14(1):17–23. <http://dx.doi.org/10.1017/S1041610202008244>. Medline:12094903
- Bossers WJR, van der Woude LHV, Boersma F, et al. The Groningen Meander Walking Test: a dynamic walking test for older adults with dementia. *Phys Ther.* 2014;94(2):262–72. <http://dx.doi.org/10.2522/ptj.20130077>. Medline:24092904
- Sterke CS, Huisman SL, van Beeck EF, et al. Is the Tinetti Performance Oriented Mobility Assessment (POMA) a feasible and valid predictor of short-term fall risk in nursing home residents with dementia? *Int Psychogeriatr.* 2010;22(2):254–63. <http://dx.doi.org/10.1017/S1041610209991347>. Medline:19951457
- Farrell MK, Rutt RA, Lusardi MM, et al. Reliability of the Physical Performance Test in people with dementia. *Phys Occup Ther Geriatr.* 2010;28(2):144–53. <http://dx.doi.org/10.3109/02703181.2010.487973>.
- Shaw FE. Prevention of falls in older people with dementia. *J Neural Transm.* 2007;114(10):1259–64. <http://dx.doi.org/10.1007/s00702-007-0741-5>. Medline:17557130
- Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. *N Engl J Med.* 1988;319(26):1701–7. <http://dx.doi.org/10.1056/NEJM198812293192604>. Medline:3205267
- Morris JC, Rubin EH, Morris EJ, et al. Senile dementia of the Alzheimer's type: an important risk factor for serious falls. *J Gerontol.* 1987;42(4):412–7. <http://dx.doi.org/10.1093/geronj/42.4.412>. Medline:3598089
- Alzheimer Society of Canada. Rising tide: the impact of dementia on Canadian society [Internet]. Toronto: Alzheimer Society of Canada; c2010 [cited 2014 May 2]. Available from: [http://www.alzheimer.ca/~media/Files/national/Advocacy/ASC\\_Rising\\_Tide\\_Full\\_Report\\_e.pdf](http://www.alzheimer.ca/~media/Files/national/Advocacy/ASC_Rising_Tide_Full_Report_e.pdf)
- Orange JB, Molloy DW, Lever JA, et al. Alzheimer's disease. Physician-patient communication. *Can Fam Physician.* 1994;40:1160–8. Medline:8019193
- Suttanon P, Hill KD, Dodd KJ, et al. Retest reliability of balance and mobility measurements in people with mild to moderate Alzheimer's disease. *Int Psychogeriatr.* 2011;23(7):1152–9. <http://dx.doi.org/10.1017/S1041610211000639>. Medline:21489342
- Tappen RM, Roach KE, Buchner D, et al. Reliability of physical performance measures in nursing home residents with Alzheimer's disease. *J Gerontol A Biol Sci Med Sci.* 1997;52(1):M52–5. <http://dx.doi.org/10.1093/gerona/52A.1.M52>. Medline:9008669
- Ries JD, Echternach JL, Nof L, et al. Test-retest reliability and minimal detectable change scores for the timed “up & go” test, the six-minute walk test, and gait speed in people with Alzheimer disease. *Phys Ther.* 2009;89(6):569–79. <http://dx.doi.org/10.2522/ptj.20080258>. Medline:19389792
- Lorbach ER, Webster KE, Menz HB, et al. Physiological falls risk assessment in older people with Alzheimer's disease. *Dement Geriatr Cogn Disord.* 2007;24(4):260–5. <http://dx.doi.org/10.1159/000107101>. Medline:17700022
- Sibley KM, Straus SE, Inness EL, et al. Balance assessment practices and use of standardized balance measures among Ontario physical therapists. *Phys Ther.* 2011;91(11):1583–91. <http://dx.doi.org/10.2522/ptj.20110063>. Medline:21868613
- Berg KO, Wood-Dauphinée SL, Williams JI, et al. Measuring balance in the elderly: preliminary development of an instrument. *Physiother Can.* 1989;41(6):304–11. <http://dx.doi.org/10.3138/ptc.41.6.304>.
- McGilton K, Wells J, Davis A, et al. Rehabilitating patients with dementia who have had a hip fracture. Part II: Cognitive symptoms that influence care. *Top Geriatr Rehab.* 2007;23(2):174–82. <http://dx.doi.org/10.1097/01.TGR.0000270186.36521.85>.
- McGilton K, Wells J, Teare G, et al. Rehabilitating patients with dementia who have had a hip fracture: Part I: Behavioural symptoms that influence care. *Top Geriatr Rehab.* 2007;23(2):161–73. <http://dx.doi.org/10.1097/01.TGR.0000270185.98402.a6>.
- Muir SW, Yohannes AM. The impact of cognitive impairment on rehabilitation outcomes in elderly patients admitted with a femoral neck fracture: a systematic review. *J Geriatr Phys Ther.* 2009;32(1):24–32. <http://dx.doi.org/10.1519/00139143-200932010-00006>. Medline:19856633
- Goldstein FC, Strasser DC, Woodard JL, et al. Functional outcome of cognitively impaired hip fracture patients on a geriatric rehabilitation unit. *J Am Geriatr Soc.* 1997;45(1):35–42. Medline:8994485
- Heruti RJ, Lusky A, Barell V, et al. Cognitive status at admission: does it affect the rehabilitation outcome of elderly patients with hip fracture? *Arch Phys Med Rehabil.* 1999;80(4):432–6. [http://dx.doi.org/10.1016/S0003-9993\(99\)90281-2](http://dx.doi.org/10.1016/S0003-9993(99)90281-2). Medline:10206606
- Seidel GK, Millis SR, Lichtenberg PA, et al. Predicting bowel and bladder continence from cognitive status in geriatric rehabilitation patients. *Arch Phys Med Rehabil.* 1994;75(5):590–3. Medline:8185455
- McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011;7(3):263–9. <http://dx.doi.org/10.1016/j.jalz.2011.03.005>. Medline:21514250
- Berg KO, Maki BE, Williams JI, et al. Clinical and laboratory measures of postural balance in an elderly population. *Arch Phys Med Rehabil.* 1992;73(11):1073–80. Medline:1444775
- Berg KO, Wood-Dauphinée SL, Williams JI, et al. Measuring balance in the elderly: validation of an instrument. *Can J Public Health.* 1992;83(Suppl 2):S7–11. Medline:1468055
- Muir SW, Berg K, Chesworth B, et al. Use of the Berg Balance Scale for predicting multiple falls in community-dwelling elderly people: a prospective study. *Phys Ther.* 2008;88(4):449–59. <http://dx.doi.org/10.2522/ptj.20070251>. Medline:18218822
- Podsiadlo D, Richardson S. The timed “up & go”: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc.* 1991;39(2):142–8. Medline:1991946
- Scott V, Votova K, Scanlan A, et al. Multifactorial and functional mobility assessment tools for fall risk among older adults in community, home-support, long-term and acute care settings. *Age Ageing.* 2007;36(2):130–9. <http://dx.doi.org/10.1093/ageing/af1165>. Medline:17293604
- Bischoff HA, Stähelin HB, Monsch AU, et al. Identifying a cut-off point for normal mobility: a comparison of the timed “up and go” test in community-dwelling and institutionalised elderly women.

- Age Ageing. 2003;32(3):315–20. <http://dx.doi.org/10.1093/ageing/32.3.315>. Medline:12720619
31. Duncan PW, Weiner DK, Chandler J, et al. Functional reach: a new clinical measure of balance. *J Gerontol*. 1990;45(6):M192–7. <http://dx.doi.org/10.1093/geronj/45.6.M192>. Medline:2229941
  32. Murphy MA, Olson SL, Protas EJ, et al. Screening for falls in community-dwelling elderly. *J Aging Phys Act*. 2003;11(1):66–80.
  33. Molloy DW, Alemayehu E, Roberts R. Reliability of a standardized Mini-Mental State Examination compared with the traditional Mini-Mental State Examination. *Am J Psychiatry*. 1991;148(1):102–5. <http://dx.doi.org/10.1176/ajp.148.1.102>. Medline:1984692
  34. American Academy of Neurology and American Geriatrics Society. Dementia: performance measurement set. Chicago: American Medical Association; 2010.
  35. Bruton A, Conway JH, Holgate ST. Reliability: what is it, and how is it measured? *Physiotherapy*. 2000;86(2):94–9. [http://dx.doi.org/10.1016/S0031-9406\(05\)61211-4](http://dx.doi.org/10.1016/S0031-9406(05)61211-4).
  36. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. New York: Oxford University Press; 2003.
  37. Stratford PW. Getting more from the literature: estimating the standard error of measurement from reliability studies. *Physiother Can*. 2005;56(1):27–30. <http://dx.doi.org/10.2310/6640.2004.15377>.
  38. Bland JM, Altman DG. Measurement in medicine: the analysis of method comparison studies. *Statistician*. 1983;32(1983):307–17. Medline:2868172
  39. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17(1):101–10. [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19980115\)17:1 < 101::AID-SIM727 > 3.0.CO;2-E](http://dx.doi.org/10.1002/(SICI)1097-0258(19980115)17:1 < 101::AID-SIM727 > 3.0.CO;2-E). Medline:9463853
  40. Holbein-Jenny MA, Billek-Sawhney B, Beckman E, et al. Balance in personal care home residents: a comparison of the Berg Balance Scale, the Multi-Directional Reach Test, and the Activities-Specific Balance Confidence Scale. *J Geriatr Phys Ther*. 2005;28(2):48–53. <http://dx.doi.org/10.1519/00139143-200508000-00003>. Medline:16236228
  41. Rockwood K, Awalt E, Carver D, et al. Feasibility and measurement properties of the functional reach and the timed up and go tests in the Canadian study of health and aging. *J Gerontol A Biol Sci Med Sci*. 2000;55(2):M70–3. <http://dx.doi.org/10.1093/gerona/55.2.M70>. Medline:10737688