BMJ Open

Inter-rater reliability of Berg Balance Scale, 30-seconds chair stand test and 6 meters walking test and construct validity of Berg Balance Scale in nursing home residents with mild and moderate dementia

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2015-008321
Article Type:	Research
Date Submitted by the Author:	26-Mar-2015
Complete List of Authors:	Telenius, Elisabeth; Oslo and Akershus University College of Applied Sciences, Health sciences, Physiotherapy Engedal , Knut ; Norwegian Centre of Aging and Health, Department of Psychiatry, Vestfold Health Trust, Bergland, Astrid; Oslo and Akershus University College of Applied Sciences, Health Sciences
Primary Subject Heading:	Rehabilitation medicine
Secondary Subject Heading:	Geriatric medicine, Neurology
Keywords:	GERIATRIC MEDICINE, Dementia < NEUROLOGY, REHABILITATION MEDICINE

Inter-rater reliability of Berg Balance Scale, 30-seconds chair stand test and 6 meters walking test and construct validity of Berg Balance Scale in nursing home residents with mild and moderate dementia

Address for correspondence: Elisabeth Wiken Telenius, Oslo and Akershus University College of Applied Sciences, PO Box 4 St Olavs plass, 0130 Oslo, Norway

Email: elisabeth-wiken.telenius@hioa.no

Telephone: +47 992 73 229

Authors:

Elisabeth Wiken Telenius PT MSc[1]; Knut Engedal MD PhD[2]; Astrid Bergland PT PhD[1]

[1] Oslo and Akershus University College of Applied Sciences, Faculty of Health Sciences, Department of Physiotherapy, Oslo, Norway

[2] Norwegian Centre of Aging and Health, Department of Psychiatry, Vestfold Health Trust, Tønsberg, Norway

Key words: Dementia, nursing home, Berg balance scale, 30 seconds chair stand test, walking speed

Word count: 2793 (excluding title page, tables and references)

Abstract

Objective: When testing physical function the patients must be alert and have capacity to understand and respond to instructions. Patients with dementia may have difficulties fulfilling these requirements and therefore the reliability of the measures may be compromised. We aimed to assess the inter-rater reliability between pairs of observers independently rating the same subject in BBS, 30 seconds chair stand test and 6 meters walking test. We also wanted to investigate the internal consistency of the BBS.

Design: Cross-sectional study

Setting: We included 33 nursing home patients with mild to moderate degree of dementia and tested them once with two evaluators present. One evaluator gave instructions and both of them scored the patients' performance. Weighted kappa and Intraclass Correlation Coefficient (ICC) model 3.1 with 95% confidence intervals were used to measure inter-rater reliability. Chronbach's alpha was calculated to evaluate the internal consistency of the BBS sum score.

Results: The mean values of the BBS scored by the two evaluators were 38 ±13.7 and 38.0 ± 13.8, respectively. Weighted kappa scores for the BBS items varied from 0.83 to 1.0. ICC for the BBS's sum score was 0.99. The Chronbach´s Alpha of BBS's sum score was 0.9. The ICC of the CST was 1 and on the 6 meters walking test it was 0.98.

Conclusion: The results reveal an excellent inter-rater reliability of the BBS, CST and 6 meter walking test as well as high internal consistency for BBS in a population of nursing home residents with mild and moderate dementia.

Article summary

- Article focus: Reliability may be compromised when testing physical function in nursing home patients with dementia. The inter-rater reliability of three commonly used physical function tests has not yet been tested in this growing population, and was therefore investigated.
- Key messages: Berg Balanse Scale, 6 meters walking test and 30 seconds Chair stand test have excellent inter-rater reliability in a population of nursing home residents with mild and moderate dementia.
- Strength and limitations of this study: The study included a well defined population with older people living in nursing home and scoring 1 or 2 on Clinical Dementia Rating Scale. Number of participants was limited.

Introduction

The worldwide prevalence of people with dementia is estimated to nearly double every 20 years, reaching 40.8 million in 2020 and 90.3 million in 2040 [1]. Dementia affects balance, mobility and gait performance [2-4], and people with dementia have a two-fold increased risk of falls compared to non-demented elderly [5]. Even though the literature is unequivocal, studies show important benefits through exercise and physical activity for older adults with dementia in areas of physical health, including activities of daily living (ADL) and of mental health [6-9]. Consequently, physical therapists are likely to be treating an increasing number of people with dementia [10]. For this reason the demand for reliable and valid measures to assess physical function in these patients will increase.[11]. According to Hauer (2008) [12], testing of physical function assumes that test participants are able to 1) comprehend the test commands, 2) develop an adequate physical action and sequence, and 3) remember both during execution of the test. Another prerequisite is that test persons show adequate attention during testing. The presence of dementia will influence these factors and could thereby affect reliability.

The lack of reliability-tested physical function instruments for patients in nursing home with dementia has been expressed repeatedly in the literature [13, 14]. To the authors´ knowledge, only one other study has investigated the reliability of BBS in a population of nursing home residents [15]. In that study 67% had dementia. They fund that the test-retest reliability was high, however, inter-rater reliability was not tested. Suttanon et al, 2011 [16] found that the reliability of different mobility and balance measures ranged between fair to excellent in a population of mostly community-dwelling elderly people with Alzheimer's disease. They stressed the importance of considering reliability when deciding which balance and mobility measures to use for this group.

Three functional tests were investigated in this study: Berg Balance Scale, 30 seconds chair stand test and 6 meters walking test. Balance is often impaired in the older persons with dementia, and improvement in balance is an important goal of rehabilitation [17]. Measuring balance can assist the clinician in selecting the most appropriate therapy and outcome measurement [18, 19]. BBS is used extensively in the clinic, has frequently been compared with other balance measures and is considered to be gold standard of measuring balance [20, 21]. Thirty seconds Chair stand test (CST) is one of the most important functional evaluation clinical tests because it measures lower body strength and relates it to the most demanding daily life activities [22, 23]. Lower limb muscle weakness has been identified as a risk factor for falls and for the inability to perform lower extremity functional tasks such as walking, sit to stand transfers, climbing steps and lower body dressing [24-26]. Walking speed is associated with reduced balance ability and increased risk of falling. It can predict health status, survival and hospital costs [27-29]. Walking speed tests are frequently used to evaluate mobility in elderly people [30, 31].

Test-retest reliability has been more frequently investigated than inter-rater reliability [10, 32]. However, during rehabilitation an elderly patient may be assessed by more than one physiotherapist, and high reliability between scorings made by different evaluators are therefore essential. This is also important when testing in multi centre research projects. We aimed to assess the inter-rater reliability between pairs of observers independently rating the same subject in BBS, CST and 6 meters walking test. We also wanted to assess the internal consistency of the BBS.

Methods

Participants:

We included 33 participants that resided in four different nursing homes in the area around Oslo, Norway. They were recruited from a randomized controlled trial that aimed to investigate the effect of a high intensity exercise program in nursing home residents with dementia. The inclusion criteria were: being above 55 years of age, having dementia of mild or moderate degree as measured by the Clinical Dementia Rating scale (CDR 1 or 2), being able to stand up alone or by the help of one person and being able to walk six meters with or without walking aid. The exclusion criteria were: patients being medically unstable, psychotic or having severe communication problems. Details about the participants can be found in table 1.

Procedure:

The study was carried out by two physiotherapists. Prior to commencing the study they had four hour practical training with the BBS. The examiners were trained in the standardised instructions of the test. The patients were tested only once in the following order: BBS, CST and 6 meter walking test, and the whole test procedure took about 30 minutes. The two physiotherapists scored the test performance simultaneously without knowledge of each other's rating ("blind") and alternated between instructing the participant and observing the patient. In this way they both administered the test in half of the patients. The reason for choosing this model was: Some of the participants were undergoing rehabilitation and could have improved, and if they had been tested on two different days within a week, their performance could have changed and thus, test-retest reliability would have been biased. Certain steps were taken to optimize the communication with the participants on all tests [33]. The progression of cuing were pre-defined and based on suggestions by Vogelpohl et al 1996 [34]. The first step was verbal cueing, which progressed to demonstrating/ mirroring and to tactile guidance and physical assistance.

Instruments:

Berg Balance Scale (BBS) is a performance-based instrument that was originally developed by Berg et al. 1989 [35] for assessment of functional balance in older adults. Berg Balance Scale (BBS) assesses performance on a 5-level scale from 0 (cannot perform) to 4 (normal performance) on 14 different tasks involving functional balance control, including transfer, turning and stepping, giving a score between zero (poor) and 56 (normal). It takes 15 to 20 minutes to complete the BBS. We used the Norwegian version of the test [36].

The 30-seconds chair stand test (CST) measures lower limb muscle strength. The score equals the number of rises from a chair in 30 seconds with arms folded across the chest [22]. During performance of the six meter walking test the participant walks six meters at comfortable speed with or without a walking aid. The time in seconds was recorded and calculated to meters per second [37].

To measure the patients' dependence/independence in the Activities of Daily Living (ADL), we employed the Barthel Index (BI), a widely used questionnaire of the activities of daily living [38, 39]. The Clinical Dementia Rating Scale (CDR), and the Mini-Mental State Examination (MMSE) were used to measure cognition. We used the CDR to validate the dementia diagnosis of the patients. Two Norwegian studies have shown that CDR staging is a valid substitute for a dementia assessment among nursing-home patients to rate dementia and dementia severity [40, 41]. The MMSE was used to assess global cognition and consists of 20 items concerning orientation, word registration and recall, attention, naming, reading, writing, following commands and figure copying [42]. Information about the participants' medical history was obtained from the medical records.

Ethics:

The study and was approved by the Regional Committee for Medical Ethics in south east of Norway 5th of September 2012. Written and verbal information about the study was given to the patients and their relatives by their primary caregiver. All the participants gave written consent to participate and were informed that they could refuse to participate at any stage in the study.

Statistics

Inter-rater reliability for the sum score of the BBS, the CST and the 6 meter walking test was measured with Intraclass Correlation Coefficients (ICCs) in SPSS version 22. The ICC quantifies the relative reliability where the relationship between two or more sets of measurements is examined. An ICC of 1 corresponds to perfect agreement. An ICC of 0.8 or higher reflects high reliability, between 0.6 and 0.8 moderate reliability and less than 0.6 indicates poor reliability [43]. Inter-rater agreement on individual items of the BBS was analysed with weighted kappa. The weighted kappa score measures the agreement among raters adjusted for the amount of agreement expected by chance and the magnitude of disagreement [44]. A kappa value of 0.75 or higher indicates excellent agreement, between 0.4 and 0.74 indicates fair to low agreement, and less than 0.4 indicates poor agreement [45]. Weighted kappa was calculated in Excel version 2011 for Mac with Real Statistics Resource Pack. Cronbach's alphas for each evaluator's scorings were calculated to assess the internal consistency of the BBS. Cronbach's Alpha is regarded as excellent when it is higher than 0.9, as good between 0.7 and 0.9 and as acceptable between 0.6 and 0.7 [46]. Internal consistency of the BBS was also tested by item-to-total correlation. An item-to-total correlation shows the degree of association between each individual item and the total score of the other items in the scale. An item-to-total correlation is considered adequate if it is above 0.4 [43].

Results

Table 1: Demographic characteristics of the participants:

Women, n (%)	25 (75.8)
Age, mean (SD)	82.7 (7.2)
Length of stay in nursing home (months), mean (SD)	22 (27.8)
Neurological disease n (%)	9 (27.3)
Heart disease n(%)	19 (57.6)
Musculoskeletal disease n(%)	9 (27.3)
MMSE-score mean (SD)	15.8 (5.4)
CDR=1 n(%)	13 (39.4%)
Barthel Index, mean (SD)	13.1 (4.4)
Walked independently n(%)	10 (30.3)
Number of diagnosis, mean (SD)	3.1 (1.8)
Number of medications, mean (SD)	6 (3.0)

Table 2: ICC of BBS, CST and 6 meters walking test

Test	Tester	Mean	Range	ICC
BBS	Tester 1	38.0	0-51	0.998
	Tester 2	38.0	0-51	
30 sec chair stand	Both testers	6 (3.2)	0-12	1
6 meter Walking test	Tester 1	0.53 (0.16)	0.22-0.84	0.98 (0.97-0.99)
	Tester 2	0.53 (0.18)	0.12-0.82	

Table 3 Distribution of BBS-scores from both evaluators: Evaluator 1 (E1) and Evaluator 2 (E2)

Items	0 point		1 point		2 points		3 points		4 points		Mean
	E1	E2	E1	E2	E1	E2	E1	E2	E1	E2	
1. Sitting to standing	4	4	0	0	0	0	10	10	19	19	3.2
2. Standing unsupported	3	3	0	0	2	2	0	0	28	28	3.5
3. Sitting unsupported	1	1	0	0	0	0	0	0	32	32	3.9
4. Standing to sitting	3	3	0	0	0	0	10	10	20	20	3.3
5. Transfers	3	3	1	1	2	1	11	13	16	15	3.1
6. Standing with eyes closed	4	3	1	1	0	1	1	1	27	27	3.4
7. Standing with feet together	8	8	2	2	5	5	5	3	13	15	2.4
8. Reaching forward with outstretched arm	5	4	2	4	10	9	14	15	2	1	2.2
9. Retrieving object from floor	5	5	2	1	0	0	0	1	26	26	3.2
10. Turing to look behind	5	5	1	0	7	8	6	7	14	13	2.7
11. Turing 360°	6	6	3	2	14	18	2	2	8	5	2.0
12. Placing alternate foot on stool	11	12	1	0	10	8	5	7	6	6	1.8
13. Standing with one foot in front	5	4	0	1	9	10	19	18	0	0	2.3
14. Standing on one foot	6	7	24	23	2	2	1	1	0	0	0.9
Total:	68	69	35	37	64	61	88	84	207	211	38.0

Table 4, Weighted kappa of the individual items of the BBS

Items	Weighted kappa between testers
1. Sitting to standing	1.00
2. Standing unsupported	1.00
3. Sitting unsupported	1.00
4. Standing to sitting	0.93
5. Transfers	0.95
6. Standing with eyes closed	0.93
7. Standing with feet together	0.96
8. Reaching forward with outstretched arm	0.87
9. Retrieving object from floor	0.89
10. Turing to look behind	0.83
11. Turing 360°	0.84
12. Placing alternate foot on stool	0.83
13. Standing with one foot in front	0.94
14. Standing on one foot	0.94
All items	0.94

Table 5: Correlation matrix

BBS	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1													
2	0.89**	1												
3	0.45**	0.51**	1											
4	0.84**	0.90**	0.51**	1										
5	0.91**	0.90**	0.45**	0.85**	1									
6	0.86**	0.81**	0.44*	0.75**	0.80**	1								
7	0.54**	0.60**	0.26	0.52**	0.54**	0.4*	1							
8	0.79**	0.64**	0.34	0.63**	0.67**	0.69**	0.44*	1						
9	0.77**	0.81**	0.37*	0.70**	0.74**	0.68**	0.51**	0.62**	1					
10	0.70**	0.69**	0.34	0.66**	0.73**	.059**	0.50**	0.63**	0.85**	1				
11	0.59**	0.54**	0.27	0.53**	0.66**	0.39*	0.50**	0.44*	0.38*	0.61**	1			
12	0.57**	0.49**	0.22	0.45**	0.60**	0.42*	0.49**	0.45**	0.63**	0.66**	0.47**	1		
13	0.84**	0.77**	0.38*	0.65**	0.79**	0.91**	0.53**	0.74**	0.66**	0.58**	0.39*	0.48**	1	
14	0.53**	0.63**	0.28	0.47**	0.55**	0.51**	0.59**	0.46**	0.57**	0.55**	0.31	0.33	0.60**	1
sum	0.93**	0.92**	0.50**	0.86**	0.93**	0.83**	0.68**	0.78**	0.86**	0.84**	0.66**	0.68**	0.84**	0.65**

Results

Demographic characteristics

Thirty-three nursing home residents (25 women, 8 men) with mild and moderate dementia participated in this study. Mean stay at the nursing home was almost two years, however it ranged between 3 months and 9 years. Four of the participants used a wheelchair, and 17 used Zimmer frames to move around. Characteristics are presented in Table 1.

Distribution of scores

The mean total score ± SD of the BBS was 38.0±3.8 for the first evaluator and 38.0±3.7 for the second evaluator (Table 2). Table 3 demonstrates the distributions on the BBS for both of the evaluators. The table shows the number of patients with a score of zero, one, two, three and four on each item. On the CST, the mean score was 6 (±3.2), ranging from 0 to 12 (table 2). The two evaluators scored identically on this test. On average the participants walked 6 meters on 12 seconds, which equals a speed of 0.5 meters per second (± 0.17 and 0.18). The results ranged between 0.22-0.84 (evaluator 1) and 0.12-0.82 (evaluator 2).

Inter-rater reliability

Weighted kappa scores for each of the 14 items on the BBS obtained by the evaluators varied from 0.83 to 1 (table 4). On the BBS, the evaluators scored differently on only 32 occasions out of the total 462, which gives an agreement percent of 93.1. Interclass correlation coefficient for the BBS's sum score was 0.998. The CST had an ICC of 1, while the 6 meters walking test ICC score was 0.98 (0.97-0.99) (table 2)

Construct validity

Cronbach's **α** coefficient of the BBS was 0.948. The correlation matrix, which included the 14 items of the BBS and sum score, are presented in Table 5. The item-to-total correlations were r >0.4 for all items except for item 3. The scores were very uniform on item 3: one participant scored 0 and the rest scored 4 points.

Discussion:

The weighted kappa in the current study ranged between 0.83 - 1, indicating an excellent inter-rater reliability when using the Berg Balance Scale in the population of nursing home residents with dementia. These results fit well with the results from other studies on other populations [35, 47-49]. The ICC of the BBS sum score was very high, which also concurs with studies on multiple sclerosis-patients [49] and people with lower limb amputations [50]. In agreement with other studies [36, 49] our findings indicate a high internal consistency of the BBS. All of the item to total correlation coefficients were 0.6 or above (except item number 3 because of little variability within scores). The high internal consistency of the BBS showed that the items of this instrument measured the same concept: balance. Some of the items show fairly high correlation, and a few correlation coefficients exceeded 0.9, which may indicate item redundancy. This should be investigated further.

In our study the mean value of BBS was 38 points. A study from three nursing homes in Sweden demonstrated a mean BBS score of 30 points [15]. The reason for this discrepancy is that our participants took part in an exercise study and therefore were more fit than the general nursing home population. On the other hand, the current population had a lower mean MMSE score (16 points), than the Swedish study (17.5 points). It is interesting to notice that even when testing a fitter group of nursing home residents, there does not seem to be a ceiling effect of BBS, as none of the participants scored the maximum amount of points on BBS [51]. Only one participant scored 0 points, which means no floor effect was detected for this population. Floor and ceiling effect have been shown in other studies [48, 52]. Our results concur with the results of Halsaa et al, 2007 [36].

The ICC of the 6 meters walking test was also very high and this has been found in similar populations by others [53]. Their study demonstrated high inter-rater reliability for both 4-meters and 6-meters walking test, with ICC of 0.96 and 0.88 in a group of elderly with cognitive impairment from both day centre and nursing home. The participants in the current study scored lower on CST (6 ± 3.2) than the similar population of Blankevoort et al: 8.1±2.95 [13]. They also had a slower walking speed, 0.5m/second ± 0.2 vs. 0.8 ± 0.3 m/second respectively. To the authors' knowledge, inter-rater reliability has never before been investigated on the CST. The two evaluators scored identically on the CST. Discrepancies in interpretation of when to not approve repetitions (participant fails to fully extend hip/ knee or does not sit down between counts) were expected, but the two evaluators agreed in all 33 performances. Both the CST and 6 meter walking test have been found to have good test-retest reliability in a similar population of elderly people with dementia living at home or in nursing home with a mean MMS score of 19 (range 10-28) [13].

Limitations of the study:
We had a relatively small sample-size; nevertheless there was sufficient information to make interesting observations in a population that is not frequently included in research studies. It is a limitation of the study that the inclusion criteria restrict our findings to nursing home residents with the ability to rise from chair with the help from one and who are able to walk 6 meters with or without walking aid. Even though some of the participants used an electrical wheel chair and managed to move 6 meters only with the help from support walkers, this means that the frailest have not been included.

Implications for practice:
This study indicates that the Berg Balance Scale, 30 seconds chair stand test and 6 meter walking test has a very good inter-rater reliability in older people with dementia living in nursing home, and the tests can be used both in research and for clinical purposes to assess physical functioning. Studies report that older persons with cognitive impairments benefit from exercise regiments [7, 54]. Our study shows that patients with mild and moderate dementia are able to take instructions, which makes reliable assessments possible.

In conclusion, our study has shown that Berg Balance Scale, 30 seconds chair stand test and 6 meters walking test have excellent inter-rater reliability in a population of nursing home residents with mild and moderate dementia.

Funding: This study is funded by Norwegian ExtraFoundation for Health and Rehabilitation.

Contributorship statement: Contribution to the design of the study: EWT, KE and AB, Drafting the work: EWT, Revising the work: KE, AB, Agree to be accountable for all aspects of the work and have approved the published version: EWT, KE, AB

Competing intrests: The authors declare no competing interests.

Data sharing statement: Data is available from the corresponding author.

References

1. Prince, M., et al., *The global prevalence of dementia: a systematic review and metaanalysis.* Alzheimers Dement, 2013. **9**(1): p. 63-75.e2.
2. van Doorn, C., A.L. Gruber-Baldini, and S. Zimmerman, *Dementia is a risk factor for falls and fall injuries among nursing home residents.* J Am Geriatr Soc, 2003. **51**: p. 1213-1218.
3. Feldman, H.H., et al., *Cognition, function, and caregiving time patterns in patients with mild-to-moderate Alzheimer disease: a 12-month analysis.* Alzheimer Dis Assoc Disord, 2005. **19**(1): p. 29-36.
4. Mazoteras Munoz, V., et al., *Gait and balance impairments in Alzheimer disease patients.* Alzheimer Dis Assoc Disord, 2010. **24**(1): p. 79-84.
5. Tinetti, M.E., et al., *Risk factors for serious injury during falls by older persons in the community.* J Am Geriatr Soc, 1995. **43**(11): p. 1214-21.
6. Rolland, Y., et al., *Exercise Program for Nursing Home Residents with Alzheimer's Disease: A 1-Year Randomized, Controlled Trial.* J Am Geriatr Soc, 2007. **55**: p. 158-165.
7. Forbes, D., et al., *Exercise programs for people with dementia.* Cochrane Database Syst Rev, 2013. **12**: p. Cd006489.
8. Thuné-Boyle, I.C.V., et al., *The effect of exercise on behavioral and psychological symptoms of dementia: towards a research agenda.* International Psychogeriatrics, 2012. **24**(7): p. 1046-1057.
9. Littbrand, H., M. Stenvall, and E. Rosendahl, *Applicability and effects of physical exercise on physical and cognitive functions and activities of daily living among people with dementia: a systematic review.* Am J Phys Med Rehabil, 2011. **90**: p. 495-518.
10. Ries, J.D., 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.* Physical Therapy, 2009. **89**(6): p. 569-579.

11. Phillips, C.D., et al., *Effects of cognitive impairment on the reliability of geriatric assessments in nursing homes.* J Am Geriatr Soc, 1993. **41**(2): p. 136-42.
12. Hauer, K. and P. Oster, *Measuring functional performance in persons with dementia.* Journal Of The American Geriatrics Society, 2008. **56**(5): p. 949-950.
13. Blankevoort, C.G., M.J. van Heuvelen, and E.J. Scherder, *Reliability of six physical performance tests in older people with dementia.* Phys Ther, 2013. **93**(1): p. 69-78.
14. Fox, B., T. Henwood, and C. Neville, *Reliability of functional performance in older people with dementia.* Australasian Journal on Ageing, 2013. **32**(4): p. 248-249.
15. Conradsson, M., et al., *Berg balance scale: intrarater test-retest reliability among older people dependent in activities of daily living and living in residential care facilities.* Phys Ther, 2007. **87**(9): p. 1155-63.
16. Suttanon, P., et al., *Retest reliability of balance and mobility measurements in people with mild to moderate Alzheimer's disease.* International Psychogeriatrics, 2011. **23**(07): p. 1152-1159.
17. Kaur, J., et al., *Rehabilitation in Alzheimer´s Disease.* Dehli Psychiatry Journal, 2013. **16**(1): p. 166-170.
18. Bohannon, R.W. and K.M. Leary, *Standing balance and function over the course of acute rehabilitation.* Arch Phys Med Rehabil, 1995. **76**(11): p. 994-6.
19. Wade, D.T., et al., *Physiotherapy intervention late after stroke and mobility.* Bmj, 1992. **304**(6827): p. 609-13.
20. Langley, F. and S. Mackintosh, *Functional balance assessment of older community dwelling adults: A systematic review of the litterature.* The internet journal of applied health sciences and practice, 2007. **5**(4).
21. Tyson, S. and L. DeSouza, *A systematic review of methods to measure balance and walking post stroke.* Physical Therapy Review, 2002. **7**(3): p. 173-186.
22. Jones, C.J., R.E. Rikli, and W.C. Beam, *A 30-s chair-stand test as a measure of lower body strength in community-residing older adults.* Res Q Exerc Sport, 1999. **70**(2): p. 113-9.
23. Millor, N., et al., *An evaluation of the 30-s chair stand test in older adults: frailty detection based on kinematic parameters from a single inertial unit.* J Neuroeng Rehabil, 2013. **10**: p. 86.
24. Azegami, M., et al., *Effect of single and multi-joint lower extremity muscle strength on the functional capacity and ADL/IADL status in Japanese community-dwelling older adults.* Nurs Health Sci, 2007. **9**(3): p. 168-76.
25. Puthoff, M.L. and D.H. Nielsen, *Relationships among impairments in lower-extremity strength and poser, functional limitations and disability in older adults.* Physical therapy, 2007. **87**: p. 1334-1347.
26. Suzuki, M., et al., *The relationship between knee extension strength and lower extremity functions in nursing home residents with dementia.* Disability and Rehabilitation, 2012. **34**(3): p. 202-209.
27. Purser, J.L., et al., *Walking speed predicts health status and hospital costs for frail elderly male veterans.* J Rehabil Res Dev, 2005. **42**(4): p. 535-46.
28. Hardy, S.E., et al., *Improvement in usual gait speed predicts better survival in older adults.* J Am Geriatr Soc, 2007. **55**(11): p. 1727-34.
29. Dumurgier, J., et al., *Slow walking speed and cardiovascular death in well functioning older adults: prospective cohort study*. Vol. 339. 2009.

30. Guralnik, J.M. and L. Ferrucci, *Assessing the building blocks of function: utilizing measures of functional limitation.* Am. J. Prev. Med, 2003. **25**: p. 112-121.
31. Wang, C.-Y., C.-J. Yeh, and M.-H. Hu, *Mobility-related performance tests to predict mobility disability at 2-year follow-up in community-dwelling older adults.* Arch. Gerontol. Geriatr, 2009.
32. Thomas, V.S. and P.A. Hageman, *A preliminary study on the reliability of physical performance measures in older day-care center clients with dementia.* International Psychogeriatrics / IPA, 2002. **14**(1): p. 17-23.
33. Chalmers, J., *Behaviour management and communication strategies for dental professionals when caring for patients with dementia.* Special Care in Dentistry, 2000. **20**(4): p. 147-154.
34. Vogelpohl, T.S., et al., *"I can do it!" Dressing: promoting independence through individualized strategies.* J Gerontol Nurs, 1996. **22**(3): p. 39-42; quiz 48.
35. Berg, K., et al., *Measuring balance in the elderly: preliminary development of an instrument.* Physiotherapy Canada, 1989. **41**(6): p. 304-311.
36. Halsaa, K.E., et al., *Assessments of Interrater Reliability and Internal Consistency of the Norwegian Version of the Berg Balance Scale.* Archives of Physical Medicine and Rehabilitation, 2007. **88**(1): p. 94-98.
37. Studenski, S., et al., *Physical performance measures in the clinical setting.* J Am Geriatr Soc, 2003. **51**(3): p. 314-22.
38. Mahoney, F.I. and D.W. Barthel, *FUNCTIONAL EVALUATION: THE BARTHEL INDEX.* Md State Med J, 1965. **14**: p. 61-5.
39. Collin, C., et al., *The Barthel ADL Index: a reliability study.* Int Disabil Stud, 1988. **10**(2): p. 61-3.
40. Nygaard, H.A. and S. Ruths, *Missing the diagnosis: senile dementia in patients admitted to nursing homes.* Scand J Prim Health Care, 2003. **21**(3): p. 148-52.
41. Engedal, K. and P.K. Haugen, *The prevalence of dementia in a sample of elderly Norwegians.* Int J Geriatr Psychiatry, 1993. **8**: p. 565-570.
42. Folstein, M.F., S.E. Folstein, and P.R. McHugh, *"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician.* J Psychiatr Res, 1975. **12**(3): p. 189-98.
43. Altman, D., *Practical statistics for medical research*. 1991, London: Chapman and Hall.
44. Cohen, J., *Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit.* Psychological Bulletin, 1968. **70**(4): p. 213-220.
45. Fleiss, J., *Statistical methods for rates and proportions*. 2nd ed. 1981, New York: John Wiley.
46. George, D. and P. Mallery, *SPSS for Windows step by step: A simple guide and reference. 11.0 update*. 4th ed. 2003, Boston: Allyn & Bacon.
47. Ottonello, M., et al., *Psychometric evaluation of the Italian version of the Berg Balance Scale in rehabilitation inpatients.* EUROPEAN JOURNAL OF PHYSICAL AND REHABILITATION MEDICINE (EUROPA MEDICOPHYSICA), 2003. **39**(4): p. 181-9.
48. Noren, A.M., et al., *Balance assessment in patients with peripheral arthritis: applicability and reliability of some clinical assessments.* Physiother Res Int, 2001. **6**(4): p. 193-204.

49. Azad, A., G. Taghizadeh, and A. Khaneghini, *Assessments of the reliability of the Iranian version of the Berg Balance Scale in patients with multiple sclerosis.* Acta Neurol Taiwan, 2011. **20**(1): p. 22-8.
50. Wong, C.K., *Interrater reliability of the Berg Balance Scale when used by clinicians of various experience levels to assess people with lower limb amputations.* Physical Therapy, 2014. **94**(3): p. 371-378.
51. Po, A.L.W., *Dictionary of Evidence-based Medicine*. 1998: Radcliffe Publishing.
52. Mao, H.F., et al., *Analysis and comparison of the psychometric properties of three balance measures for stroke patients.* Stroke, 2002. **33**(4): p. 1022-7.
53. Munoz-Mendoza, C.L., et al., *Reliability of 4-m and 6-m walking speed tests in elderly people with cognitive impairment.* Arch Gerontol Geriatr, 2011. **52**(2): p. e67-70.
54. Pitkälä, K., et al., *Efficacy of physical exercise intervention on mobility and physical functioning in older people with dementia: A systematic review.* Experimental Gerontology, 2013. **43**: p. 85-93.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2 and 3
Objectives	3	State specific objectives, including any prespecified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	4
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4 and 5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4 and 5
Bias	9	Describe any efforts to address potential sources of bias	*10*
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	NA
		(*c*) Explain how missing data were addressed	NA
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	NA
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	4
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	NA
Outcome data	15*	Report numbers of outcome events or summary measures	Table 2, 3, 4, 5
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	NA
		(*b*) Report category boundaries when continuous variables were categorized	NA
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	NA
Discussion			
Key results	18	Summarise key results with reference to study objectives	8,9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9-10
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	9-10
Generalisability	21	Discuss the generalisability (external validity) of the study results	10
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	11

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Inter-rater reliability of Berg Balance Scale, 30-seconds chair stand test and 6 meters walking test and construct validity of Berg Balance Scale in nursing home residents with mild and moderate dementia

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2015-008321.R1
Article Type:	Research
Date Submitted by the Author:	03-Jun-2015
Complete List of Authors:	Telenius, Elisabeth; Oslo and Akershus University College of Applied Sciences, Health sciences, Physiotherapy Engedal , Knut ; Norwegian Centre of Aging and Health, Department of Psychiatry, Vestfold Health Trust, Bergland, Astrid; Oslo and Akershus University College of Applied Sciences, Health Sciences
Primary Subject Heading:	Rehabilitation medicine
Secondary Subject Heading:	Geriatric medicine, Neurology
Keywords:	GERIATRIC MEDICINE, Dementia < NEUROLOGY, REHABILITATION MEDICINE

Inter-rater reliability of Berg Balance Scale, 30-seconds chair stand test and 6 meters walking test and construct validity of Berg Balance Scale in nursing home residents with mild and moderate dementia

Address for correspondence: Elisabeth Wiken Telenius, Oslo and Akershus University College of Applied Sciences, PO Box 4 St Olavs plass, 0130 Oslo, Norway

Email: elisabeth-wiken.telenius@hioa.no

Telephone: +47 992 73 229

Authors:

Elisabeth Wiken Telenius PT MSc[1]; Knut Engedal MD PhD[2]; Astrid Bergland PT PhD[1]

[1] Oslo and Akershus University College of Applied Sciences, Faculty of Health Sciences, Department of Physiotherapy, Oslo, Norway

[2] Norwegian Centre of Aging and Health, Department of Psychiatry, Vestfold Health Trust, Tønsberg, Norway

Key words: Dementia, nursing home, Berg balance scale, 30 seconds chair stand test, walking speed

Word count: 3122 (excluding title page, tables and references)

Abstract

Objective: When testing physical function the patients must be alert and have capacity to understand and respond to instructions. Patients with dementia may have difficulties fulfilling these requirements and therefore the reliability of the measures may be compromised. We aimed to assess the inter-rater reliability between pairs of observers independently rating the same subject in Berg Balance Scale (BBS), 30 seconds chair stand test (CST) and 6 meters walking test. We also wanted to investigate the internal consistency of the BBS.

Design: Cross-sectional study

Setting: We included 33 nursing home patients with mild to moderate degree of dementia and tested them once with two evaluators present. One evaluator gave instructions and both of them scored the patients' performance. Weighted kappa, Intraclass Correlation Coefficient (ICC) model 2.1 with 95% confidence intervals and minimal detectable change (MDC) were used to measure inter-rater reliability. Chronbach's alpha was calculated to evaluate the internal consistency of the BBS sum score.

Results: The mean values of the BBS scored by the two evaluators were 38 ±13.7 and 38.0 ± 13.8, respectively. Weighted kappa scores for the BBS items varied from 0.83 to 1.0. ICC for the BBS's sum score was 0.99 and the MDC 2.7 and 7%. The Chronbach´s Alpha of BBS's sum score was 0.9. The ICC of the CST and 6 meters walking test it was 1 and 0.97, respectively. The MDC on the 6 meters walking test was 0.08 and 15.2%.

Conclusion: The results reveal an excellent relative inter-rater reliability of the BBS, CST and 6 meters walking test as well as high internal consistency for BBS in a population of nursing home residents with mild and moderate dementia. The absolute reliability was 2.7 on the BBS and 0.08 on the 6 meters walking test.

Article summary

- Article focus: Reliability may be compromised when testing physical function in nursing home patients with dementia. The inter-rater reliability of three commonly used physical function tests has not yet been tested in this growing population, and was therefore investigated.

- Key messages: Berg Balanse Scale, 6 meters walking test and 30 seconds Chair stand test have excellent inter-rater reliability in a population of nursing home residents with mild and moderate dementia.

- Strength and limitations of this study: The study included a well defined population with older people living in nursing home and scoring 1 or 2 on Clinical Dementia Rating Scale. Number of participants was limited.

Introduction

The worldwide prevalence of people with dementia is estimated to nearly double every 20 years, reaching 40.8 million in 2020 and 90.3 million in 2040 [1]. Dementia affects balance, mobility and gait performance [2-4], and people with dementia have a two-fold increased risk of falls compared to non-demented elderly [5]. Even though the literature is unequivocal, studies show important benefits through exercise and physical activity for older adults with dementia in areas of physical health, including activities of daily living (ADL) and of mental health [6-9]. Consequently, physical therapists are likely to be treating an increasing number of people with dementia [10]. For this reason the demand for reliable and valid measures to assess physical function in these patients will increase.[11]. According to Hauer (2008) [12], testing of physical function assumes that test participants are able to 1) comprehend the test commands, 2) develop an adequate physical action and sequence, and 3) remember both during execution of the test. Another prerequisite is that test persons show adequate attention during testing. The presence of dementia will influence these factors and could thereby affect reliability.

The lack of reliability-tested physical function instruments for patients in nursing home with dementia has been expressed repeatedly in the literature [13, 14]. To the authors´ knowledge, only one other study has investigated the reliability of BBS in a population of nursing home residents [15]. In that study 67% had dementia. They demonstrated a high ICC-value but a relatively low absolute reliability (minimal detectable change) of 7.7 points. However, inter-rater reliability was not tested. Suttanon et al, 2011 [16] found that the reliability of different mobility and balance measures ranged between fair to excellent in a population of mostly community-dwelling elderly people with Alzheimer's disease. They stressed the importance of considering reliability when deciding which balance and mobility measures to use for this group.

Three functional tests were investigated in this study: Berg Balance Scale, 30 seconds chair stand test and 6 meters walking test. Balance is often impaired in the older persons with dementia, and improvement in balance is an important goal of rehabilitation [17]. Measuring balance can assist the clinician in selecting the most appropriate therapy and outcome measurement [18, 19]. BBS is used extensively in the clinic, has frequently been compared with other balance measures and is considered to be gold standard of measuring balance [20, 21]. The BBS has been found to have a high intra- and inter-rater reliability, but variable absolute reliability [22]. Thirty seconds Chair stand test (CST) is one of the most important functional evaluation clinical tests because it measures lower body strength and relates it to the most demanding daily life activities [23, 24]. Lower limb muscle weakness has been identified as a risk factor for falls and for the inability to perform lower extremity functional tasks such as walking, sit to stand transfers, climbing steps and lower body dressing [25-27]. Walking speed is associated with reduced balance ability and increased risk of falling. It can predict health status, survival and hospital costs [28-30]. Walking speed tests are frequently used to evaluate mobility in elderly people [31, 32].

Test-retest reliability has been more frequently investigated than inter-rater reliability [10, 33]. However, during rehabilitation an elderly patient may be assessed by more than one physiotherapist, and high reliability between scorings made by different evaluators are therefore essential. This is also important when testing in multi centre research projects. We aimed to assess the inter-rater reliability between pairs of observers independently rating the same subject in BBS, CST and 6 meters walking test. We also wanted to assess the internal consistency of the BBS.

Methods

Participants:
We included 33 participants that resided in four different nursing homes in the area around Oslo, Norway. They were recruited from a randomized controlled trial that aimed to investigate the effect of a high intensity exercise program in nursing home residents with dementia. The inclusion criteria were: being above 55 years of age, having dementia of mild or moderate degree as measured by the Clinical Dementia Rating scale (CDR 1 or 2), being able to stand up alone or with the help of one person and being able to walk six meters with or without walking aid. The exclusion criteria were: patients being medically unstable, psychotic or having severe communication problems. Details about the participants can be found in table 1.

Procedure:
The study was carried out by two physiotherapists. The examiners were trained in the standardised instructions of the tests and had experience from testing 120 patients in a study three months earlier. The patients were tested only once in the following order: BBS, CST and 6 meter walking test, and the whole test procedure took about 30 minutes. The two physiotherapists scored the test performance simultaneously without knowledge of each other's rating ("blind") and alternated between instructing the participant and observing the patient. In this way they both administered the test in half of the patients. The reason for choosing this model was: Some of the participants were undergoing rehabilitation and could have improved, and if they had been tested on two different days within a week, their performance could have changed and thus, test-retest reliability would have been biased. Certain steps were taken to optimize the communication with the participants on all tests [34]. The progression of cuing were pre-defined and based on suggestions by Vogelpohl et al 1996 [35]. The first step was verbal cueing, which progressed to demonstrating/ mirroring and to tactile guidance and physical assistance.

Instruments:
Berg Balance Scale (BBS) is a performance-based instrument that was originally developed by Berg et al. 1989 [36] for assessment of functional balance in older adults. Berg Balance Scale (BBS) assesses performance on a 5-level scale from 0 (cannot perform) to 4 (normal performance) on 14 different tasks involving functional balance control, including transfer, turning and stepping, giving a score between zero (poor) and 56 (normal). It takes 15 to 20 minutes to complete the BBS. We used the Norwegian version of the test [37].
The 30-seconds chair stand test (CST) measures lower limb muscle strength. The score equals the number of rises from a chair in 30 seconds with arms folded across the chest [23]. During performance of the six meter walking test the participant walks six meters at comfortable speed with or without a walking aid. The time in seconds was recorded and calculated to meters per second [38].

To measure the patients' dependence/independence in the Activities of Daily Living (ADL), we employed the Barthel Index (BI), a widely used questionnaire of the activities of daily living [39, 40]. The Clinical Dementia Rating Scale (CDR), and the Mini-Mental State Examination (MMSE) were used to measure cognition. We used the CDR to validate the dementia diagnosis of the patients. Two Norwegian studies have shown that CDR staging is a valid substitute for a dementia assessment among nursing-home patients to rate dementia and dementia severity [41, 42]. The MMSE was used to assess global cognition and consists of 20 items concerning orientation, word registration and recall, attention, naming, reading, writing, following commands and figure copying [43]. Information about the participants' medical history was obtained from the medical records.

Ethics:

The study and was approved by the Regional Committee for Medical Ethics in south east of Norway 5th of September 2012. Written and verbal information about the study was given to the patients and their relatives by their primary caregiver. All the participants gave written consent to participate and were informed that they could refuse to participate at any stage in the study.

Statistics

Inter-rater reliability for the sum score of the BBS, the CST and the 6 meter walking test was measured with Intraclass Correlation Coefficients (ICCs) in SPSS version 22. The ICC quantifies the relative reliability where the relationship between two or more sets of measurements is examined. An ICC of 1 corresponds to perfect agreement. An ICC of 0.8 or higher reflects high relative reliability, between 0.6 and 0.8 moderate reliability and less than 0.6 indicates poor reliability [44]. According to Shrout and Fleiss, 1979, the ICC category in the current study was case 2 because the evaluators are considered to be a random sample from a population of potential raters [45]. To test absolute reliability we calculated SEM, MDC_{95} and $MDC_{95}\%$ [46]

$SEM = SD \sqrt{(1\text{-}ICC)}$; $MDC_{95} = SEM * 1.96 * \sqrt{2}$; $MDC_{95}\% = (MDC_{95}/ \text{mean}) *100$

Inter-rater agreement on individual items of the BBS was analysed with weighted kappa. The weighted kappa score measures the agreement among raters adjusted for the amount of agreement expected by chance and the magnitude of disagreement [47]. A kappa value of 0.75 or higher indicates excellent agreement, between 0.4 and 0.74 indicates fair to low agreement, and less than 0.4 indicates poor agreement [48]. Weighted kappa was calculated in Excel version 2011 for Mac with Real Statistics Resource Pack. Cronbach's alphas for each evaluator's scorings were calculated to assess the internal consistency of the BBS. Cronbach's Alpha is regarded as excellent when it is higher than 0.9, as good between 0.7 and 0.9 and as acceptable between 0.6 and 0.7 [49]. Internal consistency of the BBS was also tested by item-to-total correlation. An item-to-total correlation shows the degree of association between each individual item and the total score of the other items in the scale. An item-to-total correlation is considered adequate if it is above 0.4 [44].

Results

Table 1: Demographic characteristics of the participants:

Women, n (%)	25 (75.8)
Age, mean (SD), range	82.7 (7.2), 66-91
Length of stay in nursing home (months), mean (SD), range	22 (27.8), 3-111
Neurological disease n (%)	9 (27.3)
Heart disease n(%)	19 (57.6)
Musculoskeletal disease n(%)	9 (27.3)
MMSE-score mean (SD), range	15.8 (5.4), 0-51
CDR=1 n(%)	13 (39.4%)
Barthel Index, mean (SD), range	13.1 (4.4), 3-20
Walked independently n(%)	10 (30.3)
Walked independently during 6 meters walking test n(%)	16 (50)
Number of diagnosis, mean (SD), range	3.1 (1.8), 1-8
Number of medications, mean (SD), range	6 (3.0), 0-13

SD: Standard deviation

Table 2: ICC of BBS, CST and 6 meters walking test

Test	Tester	Mean (SD)	Range	ICC	SEM	MDC	MDC %
BBS	Tester 1	38.0 (13.8)	0-51	0.995	0.97	1.92	7
	Tester 2	38.0 (13.7)	0-51				
30 sec chair stand	Both testers	6 (3.2)	0-12	1	0	0	0
6 meter Walking test	Tester 1	0.53 (0.16)	0.22-0.84	0.97	0.03	0.06	15.2
	Tester 2	0.53 (0.18)	0.12-0.82				

SD= Standard Deviation, ICC= Intraclass Correlation Coefficient, SEM= Standard estimate of measurement, MDC= Minimal detectable change

Table 3 Distribution of BBS-scores from both evaluators: Evaluator 1 (E1) and Evaluator 2 (E2)

Items	0 point		1 point		2 points		3 points		4 points		Mean
	E1	E2	E1	E2	E1	E2	E1	E2	E1	E2	
1. Sitting to standing	4	4	0	0	0	0	10	10	19	19	3.2
2. Standing unsupported	3	3	0	0	2	2	0	0	28	28	3.5
3. Sitting unsupported	1	1	0	0	0	0	0	0	32	32	3.9
4. Standing to sitting	3	3	0	0	0	0	10	10	20	20	3.3
5. Transfers	3	3	1	1	2	1	11	13	16	15	3.1
6. Standing with eyes closed	4	3	1	1	0	1	1	1	27	27	3.4
7. Standing with feet together	8	8	2	2	5	5	5	3	13	15	2.4
8. Reaching forward with outstretched arm	5	4	2	4	10	9	14	15	2	1	2.2
9. Retrieving object from floor	5	5	2	1	0	0	0	1	26	26	3.2
10. Turing to look behind	5	5	1	0	7	8	6	7	14	13	2.7
11. Turing 360°	6	6	3	2	14	18	2	2	8	5	2.0
12. Placing alternate foot on stool	11	12	1	0	10	8	5	7	6	6	1.8
13. Standing with one foot in front	5	4	0	1	9	10	19	18	0	0	2.3
14. Standing on one foot	6	7	24	23	2	2	1	1	0	0	0.9
Total:	68	69	35	37	64	61	88	84	207	211	38.0

Table 4, Weighted kappa of the individual items of the BBS

Items	Weighted kappa between testers
1. Sitting to standing	1.00
2. Standing unsupported	1.00
3. Sitting unsupported	1.00
4. Standing to sitting	0.93
5. Transfers	0.95
6. Standing with eyes closed	0.93
7. Standing with feet together	0.96
8. Reaching forward with outstretched arm	0.87
9. Retrieving object from floor	0.89
10. Turing to look behind	0.83
11. Turing 360°	0.84
12. Placing alternate foot on stool	0.83
13. Standing with one foot in front	0.94
14. Standing on one foot	0.94
All items	0.94

Table 5: Correlation matrix

BBS	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1													
2	0.89**	1												
3	0.45**	0.51**	1											
4	0.84**	0.90**	0.51**	1										
5	0.91**	0.90**	0.45**	0.85**	1									
6	0.86**	0.81**	0.44*	0.75**	0.80**	1								
7	0.54**	0.60**	0.26	0.52**	0.54**	0.4*	1							
8	0.79**	0.64**	0.34	0.63**	0.67**	0.69**	0.44*	1						
9	0.77**	0.81**	0.37*	0.70**	0.74**	0.68**	0.51**	0.62**	1					
10	0.70**	0.69**	0.34	0.66**	0.73**	.059**	0.50**	0.63**	0.85**	1				
11	0.59**	0.54**	0.27	0.53**	0.66**	0.39*	0.50**	0.44*	0.38*	0.61**	1			
12	0.57**	0.49**	0.22	0.45**	0.60**	0.42*	0.49**	0.45**	0.63**	0.66**	0.47**	1		
13	0.84**	0.77**	0.38*	0.65**	0.79**	0.91**	0.53**	0.74**	0.66**	0.58**	0.39*	0.48**	1	
14	0.53**	0.63**	0.28	0.47**	0.55**	0.51**	0.59**	0.46**	0.57**	0.55**	0.31	0.33	0.60**	1
sum	0.93**	0.92**	0.50**	0.86**	0.93**	0.83**	0.68**	0.78**	0.86**	0.84**	0.66**	0.68**	0.84**	0.65**

Results

Demographic characteristics

Thirty-three nursing home residents (25 women, 8 men) with mild and moderate dementia participated in this study. Mean stay at the nursing home was almost two years, however it ranged between 3 months and 9 years. Four of the participants used a wheelchair, and 17 used Zimmer frames to move around. Most common neurological diseases amongst the participants were stroke (n=3) and migraine (n=3). The most common heart diseases were hypertension (n=10), atrial fibrillation (n=4) and angina pectoris (n=3) and most common musculo-skeletal diseases were osteoporosis (n=4) and arthritis in knee or hip (n=2) Characteristics are presented in Table 1.

Distribution of scores

The mean total score ± SD of the BBS was similar between the evaluators (Table 2). Table 3 demonstrates the distributions on the BBS for both of the evaluators. The table shows the number of patients with a score of zero, one, two, three and four on each item. On the CST, the two evaluators scored identically. On average the participants walked 6 meters on 12 seconds, which equals a speed of 0.5 meters per second.

Inter-rater reliability

Weighted kappa scores for each of the 14 items on the BBS obtained by the evaluators varied from 0.83 to 1 (table 4). On the BBS, the evaluators scored differently on only 32 occasions out of the total 462, which gives an agreement percent of 93.1. Interclass correlation coefficient for the BBS's sum score was very high. The MDC indicate that a change-score of almost 3 points can be caused by the effect of being tested by a different evaluator and not necessarily clinical change. The CST had an ICC of 1, while the 6 meters walking test ICC score was 0.98 with an MDC of 0.47 (table 2)

Construct validity

Cronbach's **α** coefficient of the BBS was 0.948. The correlation matrix, which included the 14 items of the BBS and sum score, are presented in Table 5. The item-to-total correlations were r >0.4 for all items except for item 3. The scores were very uniform on item 3: one participant scored 0 and the rest scored 4 points.

Discussion:

The weighted kappa in the current study ranged between 0.83 - 1, indicating an excellent inter-rater reliability when using the Berg Balance Scale in the population of nursing home residents with dementia. These results fit well with the results from other studies on other populations [36, 50-52]. The ICC of the BBS sum score was very high, which also concurs with studies on multiple sclerosis-patients [52] and people with lower limb amputations [53]. In the current study, the MDC was 2.7, which means that one must allow for a difference in almost 3 points between evaluators. In agreement with other studies [37, 52] our findings indicate a high internal consistency of the BBS. All of the item to total correlation coefficients were 0.6 or above (except item number 3 because of little variability within scores). The high internal consistency of the BBS showed that the items of this instrument measured the same concept. Some of the items show fairly high correlation, and a few correlation coefficients exceeded 0.9, which may indicate item redundancy. This should be investigated further.

In our study the mean value of BBS was 38 points. A study from three nursing homes in Sweden demonstrated a mean BBS score of 30 points [15]. Reasons for this discrepancy may be that our participants took part in an exercise study and therefore were more fit than the general nursing home population, and that we had somewhat stricter inclusion criteria regarding physical function. On the other hand, the current population had a lower mean MMSE score (16 points), than the Swedish study (17.5 points). It is interesting to notice that even when testing a fitter group of nursing home residents, there does not seem to be a ceiling effect of BBS, as none of the participants scored the maximum amount of points on BBS [54]. Only one participant scored 0 points, which means no floor effect was detected for this population. Floor and ceiling effect have been shown in other studies [51, 55]. Our results concur with the results of Halsaa et al, 2007 [37].

The ICC of the 6 meters walking test was also very high and this has been found in similar populations by others [56]. Their study demonstrated high inter-rater reliability for both 4-meters and 6-meters walking test, with ICC of 0.96 and 0.88 in a group of elderly with cognitive impairment from both day centre and nursing home. The participants in the current study scored lower on CST (6 ± 3.2) than the similar population of Blankevoort et al: 8.1±2.95 [13]. They also had a slower walking speed, 0.5m/second ± 0.2 vs. 0.8 ± 0.3 m/second respectively. To the authors' knowledge, inter-rater reliability has never before been investigated on the CST. The two evaluators scored identically on the CST. Discrepancies in interpretation of when to not approve repetitions (participant fails to fully extend hip/ knee or does not sit down between counts) were expected, but the two evaluators agreed in all 33 performances. Both the CST and 6 meter walking test have been found to have good test-retest reliability in a similar population of elderly people with dementia living at home or in nursing home with a mean MMS score of 19 (range 10-28) [13].

Limitations of the study:

We had a relatively small sample-size; nevertheless there was sufficient information to make interesting observations in a population that is not frequently included in research studies. It is a limitation of the study that the inclusion criteria restrict our findings to nursing home residents with the ability to rise from chair with the help from one and who are able to walk 6 meters with or without walking aid. Even though some of the participants used an electrical wheel chair and managed to move 6 meters only with the help from support walkers, this means that the frailest have not been included. In the clinic there may be more than two raters, therefore it may be considered a limitation that this study only investigated the use of two evaluators. The evaluations were performed simultaneously. This may lead to an overestimation of reliability due to the fact that one evaluator watches the other evaluator instruct and score. The second evaluator may thereby gain information about the instructor's scoring through watching his/ her positioning, body language or choice of words.

Implications for practice:
This study indicates that the Berg Balance Scale, 30 seconds chair stand test and 6 meter walking test has a very good inter-rater reliability in older people with dementia living in nursing home, and the tests can be used both in research and for clinical purposes to assess physical functioning. Studies report that older persons with cognitive impairments benefit from exercise regiments [7, 57]. Our study shows that patients with mild and moderate dementia are able to take instructions, which makes reliable assessments possible.

Conclusion: The results reveal an excellent relative inter-rater reliability of the BBS, CST and 6 meters walking test as well as high internal consistency for BBS in a population of nursing home residents with mild and moderate dementia. The absolute reliability was 2.7 on the BBS and 0.08 on the 6 meters walking test.

Funding: This study is funded by Norwegian ExtraFoundation for Health and Rehabilitation.

Contribution statement: Contribution to the design of the study: EWT, KE and AB, Drafting the work: EWT, Revising the work: KE, AB, Agree to be accountable for all aspects of the work and have approved the published version: EWT, KE, AB

Competing interests: The authors declare no competing interests.

Data sharing statement: Data is available from the corresponding author. There are no unpublished data from the study.

References

1. Prince, M., et al., *The global prevalence of dementia: a systematic review and metaanalysis.* Alzheimers Dement, 2013. **9**(1): p. 63-75.e2.
2. van Doorn, C., A.L. Gruber-Baldini, and S. Zimmerman, *Dementia is a risk factor for falls and fall injuries among nursing home residents.* J Am Geriatr Soc, 2003. **51**: p. 1213-1218.
3. Feldman, H.H., et al., *Cognition, function, and caregiving time patterns in patients with mild-to-moderate Alzheimer disease: a 12-month analysis.* Alzheimer Dis Assoc Disord, 2005. **19**(1): p. 29-36.
4. Mazoteras Munoz, V., et al., *Gait and balance impairments in Alzheimer disease patients.* Alzheimer Dis Assoc Disord, 2010. **24**(1): p. 79-84.
5. Tinetti, M.E., et al., *Risk factors for serious injury during falls by older persons in the community.* J Am Geriatr Soc, 1995. **43**(11): p. 1214-21.

6. Rolland, Y., et al., *Exercise Program for Nursing Home Residents with Alzheimer's Disease: A 1-Year Randomized, Controlled Trial.* J Am Geriatr Soc, 2007. **55**: p. 158-165.
7. Forbes, D., et al., *Exercise programs for people with dementia.* Cochrane Database Syst Rev, 2013. **12**: p. Cd006489.
8. Thuné-Boyle, I.C.V., et al., *The effect of exercise on behavioral and psychological symptoms of dementia: towards a research agenda.* International Psychogeriatrics, 2012. **24**(7): p. 1046-1057.
9. Littbrand, H., M. Stenvall, and E. Rosendahl, *Applicability and effects of physical exercise on physical and cognitive functions and activities of daily living among people with dementia: a systematic review.* Am J Phys Med Rehabil, 2011. **90**: p. 495-518.
10. Ries, J.D., 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.* Physical Therapy, 2009. **89**(6): p. 569-579.
11. Phillips, C.D., et al., *Effects of cognitive impairment on the reliability of geriatric assessments in nursing homes.* J Am Geriatr Soc, 1993. **41**(2): p. 136-42.
12. Hauer, K. and P. Oster, *Measuring functional performance in persons with dementia.* Journal Of The American Geriatrics Society, 2008. **56**(5): p. 949-950.
13. Blankevoort, C.G., M.J. van Heuvelen, and E.J. Scherder, *Reliability of six physical performance tests in older people with dementia.* Phys Ther, 2013. **93**(1): p. 69-78.
14. Fox, B., T. Henwood, and C. Neville, *Reliability of functional performance in older people with dementia.* Australasian Journal on Ageing, 2013. **32**(4): p. 248-249.
15. Conradsson, M., et al., *Berg balance scale: intrarater test-retest reliability among older people dependent in activities of daily living and living in residential care facilities.* Phys Ther, 2007. **87**(9): p. 1155-63.
16. Suttanon, P., et al., *Retest reliability of balance and mobility measurements in people with mild to moderate Alzheimer's disease.* International Psychogeriatrics, 2011. **23**(07): p. 1152-1159.
17. Kaur, J., et al., *Rehabilitation in Alzheimer´s Disease.* Dehli Psychiatry Journal, 2013. **16**(1): p. 166-170.
18. Bohannon, R.W. and K.M. Leary, *Standing balance and function over the course of acute rehabilitation.* Arch Phys Med Rehabil, 1995. **76**(11): p. 994-6.
19. Wade, D.T., et al., *Physiotherapy intervention late after stroke and mobility.* Bmj, 1992. **304**(6827): p. 609-13.
20. Langley, F. and S. Mackintosh, *Functional balance assessment of older community dwelling adults: A systematic review of the litterature.* The internet journal of applied health sciences and practice, 2007. **5**(4).
21. Tyson, S. and L. DeSouza, *A systematic review of methods to measure balance and walking post stroke.* Physical Therapy Review, 2002. **7**(3): p. 173-186.
22. Downs, S., J. Marquez, and P. Chiarelli, *The Berg Balance Scale has high intra- and inter-rater reliability but absolute reliability varies across the scale: a systematic review.* J Physiother, 2013. **59**(2): p. 93-9.
23. Jones, C.J., R.E. Rikli, and W.C. Beam, *A 30-s chair-stand test as a measure of lower body strength in community-residing older adults.* Res Q Exerc Sport, 1999. **70**(2): p. 113-9.

24. Millor, N., et al., *An evaluation of the 30-s chair stand test in older adults: frailty detection based on kinematic parameters from a single inertial unit.* J Neuroeng Rehabil, 2013. **10**: p. 86.
25. Azegami, M., et al., *Effect of single and multi-joint lower extremity muscle strength on the functional capacity and ADL/IADL status in Japanese community-dwelling older adults.* Nurs Health Sci, 2007. **9**(3): p. 168-76.
26. Puthoff, M.L. and D.H. Nielsen, *Relationships among impairments in lower-extremity strength and poser, functional limitations and disability in older adults.* Physical therapy, 2007. **87**: p. 1334-1347.
27. Suzuki, M., et al., *The relationship between knee extension strength and lower extremity functions in nursing home residents with dementia.* Disability and Rehabilitation, 2012. **34**(3): p. 202-209.
28. Purser, J.L., et al., *Walking speed predicts health status and hospital costs for frail elderly male veterans.* J Rehabil Res Dev, 2005. **42**(4): p. 535-46.
29. Hardy, S.E., et al., *Improvement in usual gait speed predicts better survival in older adults.* J Am Geriatr Soc, 2007. **55**(11): p. 1727-34.
30. Dumurgier, J., et al., *Slow walking speed and cardiovascular death in well functioning older adults: prospective cohort study*. Vol. 339. 2009.
31. Guralnik, J.M. and L. Ferrucci, *Assessing the building blocks of function: utilizing measures of functional limitation.* Am. J. Prev. Med, 2003. **25**: p. 112-121.
32. Wang, C.-Y., C.-J. Yeh, and M.-H. Hu, *Mobility-related performance tests to predict mobility disability at 2-year follow-up in community-dwelling older adults.* Arch. Gerontol. Geriatr, 2009.
33. Thomas, V.S. and P.A. Hageman, *A preliminary study on the reliability of physical performance measures in older day-care center clients with dementia.* International Psychogeriatrics / IPA, 2002. **14**(1): p. 17-23.
34. Chalmers, J., *Behaviour management and communication strategies for dental professionals when caring for patients with dementia.* Special Care in Dentistry, 2000. **20**(4): p. 147-154.
35. Vogelpohl, T.S., et al., *"I can do it!" Dressing: promoting independence through individualized strategies.* J Gerontol Nurs, 1996. **22**(3): p. 39-42; quiz 48.
36. Berg, K., et al., *Measuring balance in the elderly: preliminary development of an instrument.* Physiotherapy Canada, 1989. **41**(6): p. 304-311.
37. Halsaa, K.E., et al., *Assessments of Interrater Reliability and Internal Consistency of the Norwegian Version of the Berg Balance Scale.* Archives of Physical Medicine and Rehabilitation, 2007. **88**(1): p. 94-98.
38. Studenski, S., et al., *Physical performance measures in the clinical setting.* J Am Geriatr Soc, 2003. **51**(3): p. 314-22.
39. Mahoney, F.I. and D.W. Barthel, *FUNCTIONAL EVALUATION: THE BARTHEL INDEX.* Md State Med J, 1965. **14**: p. 61-5.
40. Collin, C., et al., *The Barthel ADL Index: a reliability study.* Int Disabil Stud, 1988. **10**(2): p. 61-3.
41. Nygaard, H.A. and S. Ruths, *Missing the diagnosis: senile dementia in patients admitted to nursing homes.* Scand J Prim Health Care, 2003. **21**(3): p. 148-52.
42. Engedal, K. and P.K. Haugen, *The prevalence of dementia in a sample of elderly Norwegians.* Int J Geriatr Psychiatry, 1993. **8**: p. 565-570.

43. Folstein, M.F., S.E. Folstein, and P.R. McHugh, *"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician.* J Psychiatr Res, 1975. **12**(3): p. 189-98.
44. Altman, D., *Practical statistics for medical research*. 1991, London: Chapman and Hall.
45. Shrout, P.E. and J.L. Fleiss, *Intraclass correlations: uses in assessing rater reliability.* Psychol Bull, 1979. **86**(2): p. 420-8.
46. Stratford, P., *Getting more from the literature: estimating the standard error of measurement from reliability studies.* Physiother Can, 2004. **56**: p. 27-30.
47. Cohen, J., *Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit.* Psychological Bulletin, 1968. **70**(4): p. 213-220.
48. Fleiss, J., *Statistical methods for rates and proportions*. 2nd ed. 1981, New York: John Wiley.
49. George, D. and P. Mallery, *SPSS for Windows step by step: A simple guide and reference. 11.0 update*. 4th ed. 2003, Boston: Allyn & Bacon.
50. Ottonello, M., et al., *Psychometric evaluation of the Italian version of the Berg Balance Scale in rehabilitation inpatients.* EUROPEAN JOURNAL OF PHYSICAL AND REHABILITATION MEDICINE (EUROPA MEDICOPHYSICA), 2003. **39**(4): p. 181-9.
51. Noren, A.M., et al., *Balance assessment in patients with peripheral arthritis: applicability and reliability of some clinical assessments.* Physiother Res Int, 2001. **6**(4): p. 193-204.
52. Azad, A., G. Taghizadeh, and A. Khaneghini, *Assessments of the reliability of the Iranian version of the Berg Balance Scale in patients with multiple sclerosis.* Acta Neurol Taiwan, 2011. **20**(1): p. 22-8.
53. Wong, C.K., *Interrater reliability of the Berg Balance Scale when used by clinicians of various experience levels to assess people with lower limb amputations.* Physical Therapy, 2014. **94**(3): p. 371-378.
54. Po, A.L.W., *Dictionary of Evidence-based Medicine*. 1998: Radcliffe Publishing.
55. Mao, H.F., et al., *Analysis and comparison of the psychometric properties of three balance measures for stroke patients.* Stroke, 2002. **33**(4): p. 1022-7.
56. Munoz-Mendoza, C.L., et al., *Reliability of 4-m and 6-m walking speed tests in elderly people with cognitive impairment.* Arch Gerontol Geriatr, 2011. **52**(2): p. e67-70.
57. Pitkälä, K., et al., *Efficacy of physical exercise intervention on mobility and physical functioning in older people with dementia: A systematic review.* Experimental Gerontology, 2013. **43**: p. 85-93.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2 and 3
Objectives	3	State specific objectives, including any prespecified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	4
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4 and 5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4 and 5
Bias	9	Describe any efforts to address potential sources of bias	*10*
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	NA
		(*c*) Explain how missing data were addressed	NA
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	NA
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	4
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	NA
Outcome data	15*	Report numbers of outcome events or summary measures	Table 2, 3, 4, 5
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	NA
		(*b*) Report category boundaries when continuous variables were categorized	NA
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	NA
Discussion			
Key results	18	Summarise key results with reference to study objectives	8,9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9-10
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	9-10
Generalisability	21	Discuss the generalisability (external validity) of the study results	10
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	11

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.