International Psychogeriatrics (2011), 23:7, 1152–1159 © International Psychogeriatric Association 2011
doi:10.1017/S1041610211000639

Retest reliability of balance and mobility measurements in people with mild to moderate Alzheimer's disease

Plaiwan Suttanon,[1,2] Keith D. Hill,[2,3,4] Karen J. Dodd[1,3] and Catherine M. Said[5,6]

[1] *School of Physiotherapy, Faculty of Health Sciences, La Trobe University, Bundoora, Victoria, Australia*
[2] *Preventive and Public Health Division, National Ageing Research Institute, The Royal Melbourne Hospital, Parkville, Victoria, Australia*
[3] *Musculoskeletal Research Centre, Faculty of Health Sciences, La Trobe University, Bundoora, Victoria, Australia*
[4] *Allied Health Division, Northern Health, Bundoora, Victoria, Australia*
[5] *Physiotherapy, School of Health Science, The University of Melbourne, Parkville, Victoria , Australia*
[6] *Physiotherapy Department, Heidelberg Repatriation Hospital, Heidelberg West, Victoria Australia*

ABSTRACT

Background: To interpret changes of balance and mobility in people with Alzheimer's disease (AD), we require measures of balance and mobility that have demonstrated reliability in this population. The aim of the study was to determine the safety, feasibility and retest reliability of clinical and forceplate balance and mobility measurements in people with AD.

Methods: Relative and absolute reliabilities were examined in 14 older people with mild to moderate AD. Relative reliability was calculated using the intraclass correlation coefficient, two-way mixed model ($ICC_{3,1}$). Absolute reliability was calculated using the standard error of measurement (SEM), the minimum detectable change (MDC) and the coefficient of variation (CV).

Results: All measurements were clinically feasible and could be safely administered. ICC values were excellent and CVs were less than 11% in all clinical balance and mobility measures except the Timed Up & Go test with cognitive or manual task ($ICC_{3,1} = 0.5$ and 0.7, and CV = 14% and 10%, respectively). Most balance and mobility measures tested on the Neurocom^{TM} forceplate (modified Clinical Test of Sensory Interaction on Balance, Walk Across (step width, step length parameters), and Sit to Stand (rising index parameter)) had excellent relative reliability ($ICC_{3,1}$ ranging from 0.75 to 0.91). ICC values were fair to good for the other measures.

Conclusions: Retest reliability of the balance and mobility measures used in this study ranged between fair to good, and good to excellent. Clinicians should consider retest reliability when deciding which balance and mobility measures are used to assess people with AD.

Key words: postural control, mobility, assessment, reliability, Alzheimer's disease

Introduction

Older people with Alzheimer's disease (AD) have greater balance and mobility impairments, and corresponding falls risk, compared with their cognitively intact peers (Manckoundia *et al.*, 2006; Rolland *et al.*, 2009). For clinicians to be able to interpret changes in balance and mobility in people with AD in response to interventions, measures of balance and mobility with demonstrated retest reliability in people with AD are required. It is also important that measures of balance and mobility used to assess populations with increased risk such as those with AD are safe and feasible to implement.

Reliability can be either relative or absolute (Atkinson and Nevill, 1998). Relative reliability provides an estimate of the ability of a measurement procedure to differentiate among people (Atkinson and Nevill, 1998), rather than an estimate of measurement consistency (Stratford, 1989). Correlation coefficients are commonly used to determine relative reliability. Absolute reliability provides an estimate of measurement consistency and this type of reliability provides a reference threshold for the amount of change required for a statistically significant difference in a repeated measure (Stratford, 1989; Bruton *et al.*, 2000). The standard error of measurement (SEM), the minimum detectable change (MDC) and the coefficient of variation

Correspondence should be addressed to: Plaiwan Suttanon, PO Box 2127, Royal Melbourne Hospital (Royal Park Campus), VIC 3050, Australia. Phone: +61 3 8387 2383, Fax: +61 3 8387 2153. Email: p.suttanon@nari.unimelb.edu.au; psuttanon@students.latrobe.edu.au. Received 17 Dec 2010; revision requested 23 Feb 2011; revised version received 28 Feb 2011; accepted 3 Mar 2011. First published online 14 April 2011.

(CV) are common ways of evaluating absolute reliability. To fully evaluate retest reliability, relative and absolute estimates of retest reliability should be used together (Bruton *et al.*, 2000).

Most measures of balance and mobility tested in older cognitively intact people have good relative and absolute retest reliability. However, measures of reliability are specific to the populations and testing procedures used (Phillips *et al.*, 1993), therefore results of studies with cognitively intact older people may not apply to people with AD. In particular, measures with complex instructions or tasks may have lower reliability in people with cognitive impairment (Tappen *et al.*, 1997).

Few previous studies have examined retest reliability of balance and mobility measures in people with cognitive impairment. One such study (van Iersel *et al.*, 2007) of 39 people with dementia reported moderate to high retest reliability (ICCs = 0.56–0.97) of the Berg Balance Scale, Tinetti gait and balance test score, the Timed Up & Go (TUG) and walking tests over two weeks and sensitivity to change (responsiveness index 0.4 to 7.2). Another study of 12 people with dementia reported high retest reliability (ICC = 0.87) of the TUG over one week, although a quarter of participants were unable to perform the test (Thomas and Hageman, 2002). While these studies suggest that relative retest reliability might be adequate for people with dementia, a limitation of these two previous studies is that estimates of absolute reliability were not reported. In addition, the types of dementia were not specified. Knowing dementia type is important because cognitive and physical performance differs between dementia types (Mitnitski *et al.*, 1999), and so reliability may also differ.

Of particular interest is the reliability of balance and mobility measures in people with AD because this is the most common condition causing dementia. In a study of 33 institutionalized people with AD the retest reliability of the Timed Stand and Walk test (modified TUG) and walking tests one week apart was moderate to high (ICCs = 0.50–0.84) (Tappen *et al.*, 1997). However, because of the severity of participant symptoms (later stages of AD) test modifications were required, and so results may not be applicable to tests using standard procedures. In another study of 20 people with mild to moderate AD, ICCs of walking tests measured one week apart were high (ICC > 0.86) and CV and MDC values were good (Wittwer *et al.*, 2008). However, a limitation of this study was that it focused on gait measurement and did not include more complex balance and mobility tasks. Neither study reported the feasibility or safety of the procedures for people with AD (Tappen *et al.*, 1997; Wittwer *et al.*, 2008).

Forceplate technology has also been used to demonstrate impaired balance and mobility in people with mild to moderately severe AD (Chong *et al.*, 1999; Manckoundia *et al.*, 2006; Leandri *et al.*, 2009). Unfortunately, the safety, feasibility and retest reliability of forceplate measures of balance and mobility have not been investigated in people with AD. This is a limitation because information about the retest reliability of clinical and forceplate balance and mobility measures would facilitate the interpretation of results and the safety and feasibility would assist in judging the suitability of these measures for research and clinical use with people with AD.

Given these considerations, the primary purpose of this study was to examine the absolute and relative retest reliability of commonly used clinical and instrumental (forceplate) measures for balance and mobility performance in people with AD. The secondary purpose of the study was to determine the safety and feasibility of these tests.

Methods

Participants

The participants were a subsample from an ongoing study investigating the effectiveness of an exercise program in people with AD, who had agreed to undergo a repeat assessment one week after this scheduled baseline or post intervention assessment for that study. Volunteers for this study were at least 65 years old, had a diagnosis of AD (confirmed by a medical practitioner) of mild to moderate severity (Mini-Mental State Examination score ≥ 10), were independently mobile, and lived in the community. Of 17 participants from the exercise study who were approached to participate in this reliability study, 14 agreed to participate (82% response rate). Main reasons for declining participation were related to not wanting to do another detailed assessment. The study was approved by the relevant Human Ethics Committees and written informed consent was obtained from participants or their carers.

Procedure

Participant profile information was collected which included the Mini-Mental State Examination (MMSE; Folstein *et al.*, 1975), the Frontal Assessment Battery (FAB; Dubois *et al.*, 2000), medical conditions and medications, and history of falls in the preceding 12 months (verified by carer).

Participants then completed an assessment battery of balance and mobility tests on two occasions held approximately one week apart. Due to the complexity of the balance system (Huxham *et al.*, 2001), a multidimensional framework was used to select measures to provide a broad range of domains related to global balance performance, including static and dynamic balance, maintaining stability during functional tasks, and single and dual task performance. A single assessor with six years of physiotherapy experience conducted all measurements. To assess feasibility, the percentage of participants who could perform the tests each session was recorded. To assess safety, issues such as a slip or fall during testing were systematically recorded.

Participants were given simple, clear instructions. To help participants remember the full task, verbal cues were given during each test, and demonstrations were performed either before or during the tests (if applicable). Each participant was allowed to practice each task prior to data collection on each test occasion. Participants were tested in bare feet to remove any performance variability associated with different shoes (Menant *et al.*, 2008). The detailed test procedure including instructions and verbal cues for each measurement is available from the authors. All measurements were conducted at the Gait and Balance Laboratory of the National Ageing Research Institute.

The Functional Reach (FR) test assessed dynamic bilateral stance balance (Duncan *et al.*, 1990). The participant stood alongside (but not leaning against) a wall with feet 10 cm apart, their dominant arm raised to 90 degrees. The participant was instructed to reach forward as far as possible (FR score was the additional reach from the starting position in cm).

The Step Test (ST) assessed dynamic single leg standing balance (Hill *et al.*, 1996). The participant stood with feet parallel and 10 cm apart, with a 7.5 cm block placed 5 cm directly in front. The participant was instructed to place the whole foot onto the block, then return it fully back down to the floor repeatedly as fast as possible for 15 seconds. Each leg was tested separately, and performance on the left or right side with the least number of steps was used for data analysis.

The Timed Up & Go (TUG) test (Podsiadlo and Richardson, 1991) required the participant to stand up from a standard chair, walk 3 meters at their usual pace, turn, walk back and sit down again in the chair (timed in seconds). This task was reassessed under two dual task conditions: (1) TUG with manual task, where the participant completed the TUG (as described above) while carrying a full cup of water; and (2) TUG with cognitive task, where the participant completed the TUG while counting backwards by 3s from a randomly selected number between 20 and 100 (Shumway-Cook *et al.*, 2000).

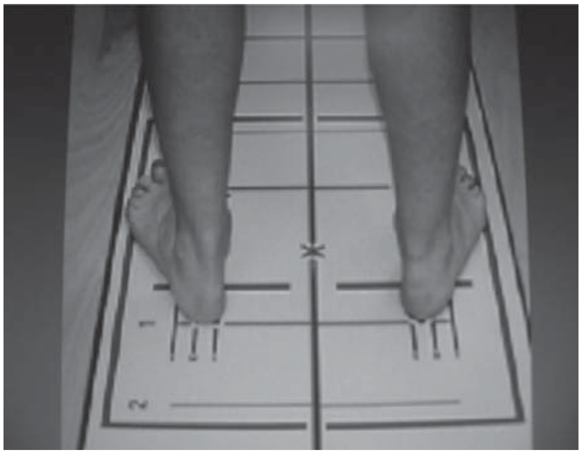

Figure 1. Feet position during the modified Clinical Test of Sensory Interaction on Balance (mCTSIB) and the Limits of Stability test (LOS) on the NeuroCom Balance Master^{TM}.

The modified Clinical Test of Sensory Interaction on Balance (mCTSIB) assessed static balance under four sensory conditions (eyes open and eyes closed standing on a firm surface, and eyes open and eyes closed while standing on foam). A safety harness was worn to prevent a fall, but did not constrain trunk movement during normal balance adjustments. Each condition consisted of three trials of 10 seconds, assessed on the NeuroCom Balance Master^{TM} (long plate). The participant's feet were positioned apart at one of three standardized foot positions related to the participant's height according to the manufacturer's instructions (NeuroCom^{TM} International Inc., 2003) (Figure 1). The assessor stood beside the platform, and if a carer was present, the carer was asked to stand on the other side of the platform, to reassure the participant that they would be safe. Instructions were to look straight ahead and stand as still as possible. Sway velocity (degrees/second) in each condition was measured, and a summary composite score was used for data analysis.

The Limits of Stability (LOS) was assessed using the NeuroCom Balance Master^{TM} (long plate) and measured participants' ability to move their center of gravity (COG) by shifting their weight to eight targets positioned in an ellipse, with the perimeter at 100% of their theoretic limits of stability (computed by the NeuroCom^{TM} system, based on the participant's height) (Figure 2). The test was performed with the same feet position as the mCTSIB, and the safety harness was worn. Instructions were to move the cursor on the screen from the centre box directly to each target box (one of the eight targets highlighted on the screen) as fast and as close to the target as possible and then hold steady for 8 seconds. A practice was allowed prior to the actual test for each direction. Measures included Reaction Time (seconds), Movement Velocity (degree/second), Maximum Excursion (%

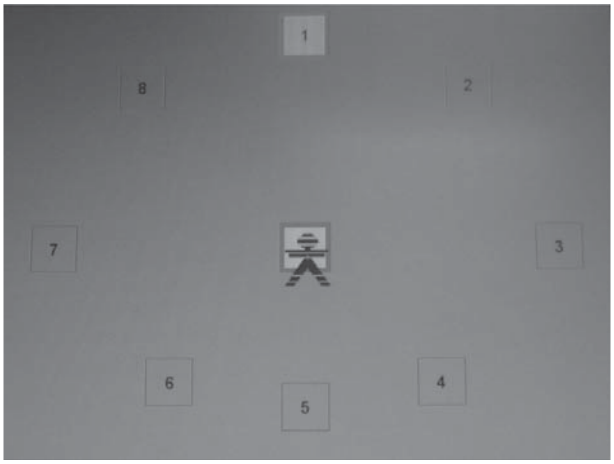

Figure 2. One centered box and eight target boxes positioned at the 100% of the Limits of Stability in the Limits of Stability test (LOS) on the NeuroCom Balance Master^{TM}. The cursor (indicating the person's center of pressure) is the stick figure.

of LOS boundary) and Directional Control (%) in each of the eight directions. A summary composite score integrating data from each of the eight directions assessed, computed by the NeuroCom^{TM} system for each measure (described above), was used for analysis.

The Walk Across test assessed gait characteristics, including step width (cm), step length (cm), and walking speed (cm/second) as the participant walked at a comfortable speed across the long plate (NeuroCom Balance Master^{TM}). To start the test, the participant was positioned approximately 90 cm in front of the platform, on a surface level with the platform. Three trials were conducted and the average score was used for analysis.

The Step Quick Turn test measured stability during turning on the NeuroCom Balance Master^{TM} forceplate. The test required the participant to take two steps forward, quickly turn and return to the starting point. To start the test, participants were positioned at the front of the forceplate. Performance was assessed turning separately to the left and to the right, with three trials in each direction. The participant was advised to use their preferred method of turning. Average measures of turn time (seconds) and turn sway (degrees/second) were derived for turning to the right and turning to the left. Data for the turn direction with worst (highest) score for turn time and turn sway were used in analyses.

The Timed Chair Stand test assessed functional leg muscle strength (Whitney *et al.*, 2005). Each participant was asked to sit at the front edge of a 45 cm-high chair with their arms crossed, and instructed to stand up fully and sit down five times as quickly as possible. The number of sit-to-stand repetitions was counted aloud to ensure participants were aware of the number of repetitions completed and to encourage them to complete the task. Time to complete the task was recorded (seconds).

Sit to Stand (on the NeuroCom Balance Master^{TM}) was used to assess stability during standing up from a seated position without upper extremity assistance. The participant sat on a 40 cm-high block, positioned centrally on the forceplate. The test was performed three times on the forceplate. Measures of the force exerted by the legs during the rising phase (rising index), expressed as% body weight, and COG sway velocity (degrees/second) were reported.

Statistical analyses

The required sample size was calculated using the method described by Walter *et al.* (1998). The few studies investigating retest reliability of physical performance measures in people with dementia or AD have reported reliability coefficients in the range of 0.6 and 0.9 (Thomas and Hageman, 2002; Lorbach *et al.*, 2007). The formula of Walter *et al.* (1998) for sample size estimate for intra-rater reliability studies requires an estimate of the desired / preferred reliability (informed by other studies if data are available), and an estimate of the lowest acceptable level of reliability. Using this formula and adopting an optimal/desired reliability level of ICC = 0.9, and a lowest acceptable reliability of ICC = 0.65, the sample size estimate indicated 14 participants were required.

Relative reliability was evaluated using ICCs, two-way mixed model ($ICC_{3,1}$). ICCs were interpreted as excellent (values > 0.75); fair to good (ICCs 0.40–0.75); and poor (ICCs < 0.40) (Shrout and Fleiss, 1979).

Three indices of absolute reliability were used. First, the standard error of measurement (SEM) which expresses measurement error in the same units as the measurement itself was calculated using the mean square error (MSE) from a repeated measure ANOVA, where $SEM = \sqrt{MSE}$ (Stratford and Goldsmith, 1997; Atkinson and Nevill, 1998). Second, the minimum detectable change (MDC_{95}), an estimate of the smallest change in score that can be detected beyond measurement error, was calculated for each test ($MDC_{95} = SEM \times \sqrt{2} \times$ the z score of 1.96) (Standford, 2004). Third, the coefficient of variation (CV), a measure of variability that is expressed as a dimensionless coefficient, was calculated. This allows the reliability of different measurement procedures to be compared (Atkinson and Nevill, 1998). The CV was calculated based on the mean square error term of logarithmically transformed data as described by Hopkins (2000).

Table 1. Participants' characteristics

CHARACTERISTICS	VALUES
Age (years)	79.57 ± 6.19
Sex (male : female)	7:7
Height (cm)	163.75 ± 12.51
Living arrangements	
- Home, no carer – n (%)	3 (21.4)
- Home, with carer – n (%)	11 (78.6)
Most common co-morbidities	
- Arthritis / osteoporosis – n (%)	6 (42.9)
- Lower limb joint replacement – n (%)	3 (21.4)
- Back pain – n (%)	4 (28.6)
- Hypertension – n (%)	8 (57.1)
- Diabetes Mellitus – n (%)	5 (35.7)
Number of prescription medications	6.29 ± 3.10
Weight (kg)	66.12 ± 13.80
MMSE	21.43 ± 5.00
FAB	13.14 ± 2.35
Number of falls in the past 12 months	0.93 ± 1.98
Number of fallers in the preceding year – n (%)	4 (28.6)
Number of multiple fallers in the preceding year – n (%)	2 (14.3)
Test-retest interval (days)	8.21 ± 3.21

Values are mean ± standard deviation (SD), or as otherwise indicated.
MMSE = Mini-Mental State Examination; FAB = Frontal Assessment Battery.

Statistical analyses were performed with SPSS Graduate Student Version 16.0 for Windows, and Microsoft Excel.

Results

Fourteen participants (50% female; mean age 79.7 years) were recruited; primarily through memory clinics. Sample demographics are summarized in Table 1.

All measurements included in this study could be administered safely to participants with AD. However, on one or both test occasions, 30% of participants were incapable of performing the Timed Up & Go test with the cognitive task. No participants had any falls or slips during testing. Table 2 presents the percentage of participants who could complete each test, the ICC values, and the three absolute reliability values. As the table shows, ICC retest estimates were excellent (above 0.75) for most clinical measures (Functional Reach, Step Test, Timed Chair Stand and Timed Up & Go (TUG) test-single task). ICCs for the TUG test assessed under dual task conditions (manual and cognitive) indicated fair to good reliability (0.40–0.75). The CVs were less than 10% for Functional Reach and TUG test; but slightly larger for Step Test, Timed Chair Stand, the TUG with a secondary manual task and TUG with a cognitive task, where the CVs ranged from 10.5% to 14.1%.

ICC values for three of the balance and mobility measures tested on the NeuroCom^{TM} forceplate [the modified Clinical Test of Sensory Interaction on Balance, Walk Across (for step width and step length parameters), and Sit to Stand (rising index parameter)] ranged from 0.75 to 0.91, indicating excellent reliability. The CVs for these measures were approximately 14%, with the exception of the Sit to Stand test (rising index parameter, CV = 7.7%). For the Limits of Stability (LOS) test, ICC values for all parameters ranged from 0.48 to 0.71 and the CVs ranged from 6.2% to 14.7%. ICCs for the Step Quick Turn measures (turn sway and turn time) were fair to good, and CVs were 10.5% and 14.4%.

MDC_{95} values are also reported in Table 2. These provide the range of scores considered to be within the range of measurement error, in the units of the actual measurement. For example, using the Step Test in an intervention study with participants with AD, a change of more than 2.42 steps would be required to be confident that the difference reflected true change.

Discussion

Generally, the clinical measures of balance and mobility included in the study showed excellent retest reliability. Given the multidimensional nature of balance performance, these results, in combination with previous studies that have demonstrated high retest reliability for the Berg Balance Scale and the TUG in people with dementia (Thomas and Hageman, 2002; van Iersel *et al.*, 2007), give clinicians confidence that these measures can be used to evaluate changes in response to interventions or disease progression over time in this clinical group. More complex clinical measures, such as the dual task TUG, had lower levels of reliability. This could in part be due to the impaired ability of participants to process the more complex instructions associated with these tasks.

Overall, assessments using the NeuroCom^{TM} forceplate showed good reliability, with excellent ICCs for the modified Clinical Test of Sensory Integration on Balance, the Walk Across (step width and step length), and Sit to Stand (rising index variable) measures. Although conducted on the forceplate, these measures incorporate ordinary tasks (e.g. standing and walking) into the assessment. This familiarity could partially explain their higher level of reliability compared

Table 2. Retest reliability indices for the tests of balance and mobility

OUTCOME MEASURES (n,%)	MEAN TEST ± SD	MEAN RETEST ± SD	ICC3,1	SEM	CV (%)	MDC$_{95}$
Static Balance (tested on forceplate)						
- mCTSIB_composite score (deg/sec), (14,100%)	1.28 ± 0.56	1.30 ± 0.54	0.911	0.17	14.9	0.34
Dynamic Balance (clinical measures)						
- Functional Reach (cm), (14, 100%)	28.98 ± 4.40	28.95 ± 3.32	0.840	1.61	5.7	3.15
- Step Test_worst leg (step) (14, 100%)	13.21 ± 2.99	13.07 ± 3.58	0.868	1.24	11.3	2.42
Dynamic Balance (tested on forceplate)						
- LOS_reaction time (sec), (14, 100%)	1.06 ± 0.15	1.13 ± 0.26	0.515	0.15	14.2	0.29
- LOS_movement velocity (deg/sec), (14, 100%)	3.11 ± 0.68	3.44 ± 0.67	0.480	0.46	14.7	0.91
- LOS_maximum excursion (%LOS boundary), (14, 100%)	75.14 ± 10.11	80.07 ± 8.06	0.677	4.44	6.2	8.71
- LOS_directional control (%), (14, 100%)	66.21 ± 10.91	66.86 ± 7.95	0.713	5.24	8.3	10.27
Mobility (clinical measures)						
- Timed Up & Go (sec), (14, 100%)	13.88 ± 2.95	13.73 ± 1.80	0.757	1.24	9.4	2.42
- TUG with manual task (sec), (14, 100%)	15.09 ± 2.81	15.04 ± 2.31	0.700	1.45	10.1	2.83
- TUG with cognitive task (sec), (10, 71.4%)	18.47 ± 3.01	18.12 ± 3.65	0.512	2.39	14.1	4.69
Mobility (tested on forceplate)						
- Walk Across – step width (cm/ sec), (14, 100%)	15.44 ± 4.25	16.24 ± 3.62	0.886	1.26	14.7	2.48
- Walk Across – step length (cm), (14, 100%)	35.59 ± 8.99	35.81 ± 8.91	0.751	4.59	13.9	9.00
- Walk Across – speed (cm), (14, 100%)	42.74 ± 10.02	44.31 ± 11.03	0.495	7.58	20.6	14.86
- Step Quick Turn – time (sec), (14, 100%)	2.52 ± 0.48	2.28 ± 0.53	0.545	0.33	14.4	0.64
- Step Quick Turn – sway (degree), (14, 100%)	42.71 ± 7.65	42.41 ± 7.12	0.637	4.56	10.5	8.93
Strength related balance performance						
- Timed Chair Stand (sec), (14, 100%)	12.49 ± 2.81	13.28 ± 3.55	0.797	1.39	10.5	2.73
- Sit to Stand – rising index (% body weight), (14, 100%)	15.79 ± 5.48	16.00 ± 5.48	0.951	1.25	7.7	2.44
- Sit to Stand – sway velocity (deg/ sec), (14, 100%)	3.86 ± 1.11	4.44 ± 1.30	0.019	1.20	39.2	2.35

mCTSIB = modified Clinical Test of Sensory Interaction on Balance; LOS = Limits of Stability test; TUG = Timed Up & Go test.

with some of the more complex forceplate measures. The remaining forceplate measures demonstrated poor (Sit to Stand sway) or only fair to good ICCs (LOS velocity, reaction time and maximum excursion; Walk Across speed; and the Step Quick Turn sway and turn time measures). The CVs were substantially larger as well for most of these measures. Several of these tasks do seem to require the ability to plan a movement sequence, which is likely to be a factor in the lower reliability for these tasks. As such, use of these measures may need to be considered carefully in people with AD, as substantially larger effect or group differences will need to be evident for significant changes to be observed using these measures compared to measures with higher ICCs and lower SEMs, CVs and MDCs.

Almost all of the measures with good to excellent ICCs also revealed acceptable levels of the CVs ($\leq$ 11%). In two forceplate measures – the modified CTSIB and Walk Across (step width) – the ICCs were excellent but had relatively high CVs (approximately 14%) indicating moderate differences in the repeated measures compared with their mean scores. The implications of this are that these measurements could only reliably differentiate moderate to large changes in performance. The relatively high MDC of the modified CTSIB suggests that large changes are required to ensure actual clinical change in people with AD.

The absolute reliability estimates for balance and mobility measures have rarely been reported in people with AD. However, one previous study investigated the absolute reliability (CV, MDC) of gait analysis using an instrumented walkway (GAITRite® system) in 20 people with mild to moderate AD (Wittwer *et al.*, 2008). The CVs and MDCs of the gait measures in our study are generally higher suggesting poorer reliability compared with the CV and MDC values of the previous study. This is despite the fact that they were derived from a similar sample. It is possible that the shorter walkway in this study (360 cm compared with the 830 cm walkway in the previous study) meant that the acceleration and deceleration phases of walking were included in the gait analysis. This may have impacted on the reliability. Therefore, for assessment of gait measures in research and clinical practice, we recommend using a longer walkway to measure only steady state walking.

Even though we found good reliability for most clinical measures, and an acceptable level of reliability for the force platform balance measures, these derived from a relatively small sample and this could possibly limit the generalizability of the results. Further research with larger samples would be able to determine the reliability estimates with greater precision. A number of factors influence a clinician's choice of which measures of balance and mobility to use for specific clinical populations. Retest reliability is an important consideration. Using the procedures reported in this study, those measures with lower levels of retest reliability should be considered with some caution for people with AD. However, retest reliability for these measures may be able to be improved with additional levels of practice, or different instructions being used.

Our results indicate that the clinical and forceplate measures of balance and mobility included in the current study were safely undertaken by people with mild to moderate AD. The Timed Up & Go test with cognitive task proved the least feasible measure, with four out of 14 participants not being able to complete the test on either test occasion. This might suggest that selecting an easier cognitive task such as reciting backwards the months of the year may be more applicable for this population.

Conclusions

Given additional cues and appropriate demonstration, all except one of the clinical and instrumental balance and mobility measures reported in this study are feasible and can be safely used to assess people with AD with fair to good or good to excellent retest reliability.

Conflict of interest

None.

Description of author's roles

All authors contributed to the study concept and design. Data were collected by P. Suttanon under K. D. Hill's supervision. Statistical analysis and interpretation of data were carried out by K. D. Hill and P. Suttanon. Drafting of the manuscript was conducted by P. Suttanon, with critical revision of the manuscript by all other authors.

Acknowledgments

We thank the people with AD and their carers who participated in this study.

References

Atkinson, G. and Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, 26, 217–238.

Bruton, A., Conway, J. H. and Holgate, S. T. (2000). Reliability: what is it, and how is it measured? *Physiotherapy*, 86, 94–99.

Chong, R. K. Y., Horak, F. B., Frank, J. and Kaye, J. (1999). Sensory organization for balance: specific deficits in Alzheimer's but not in Parkinson's disease. *Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, 54, M122–M128.

Dubois, B., Slachevsky, A., Litvan, I., and Pillon, B. (2000). The FAB: a frontal assessment battery at bedside. *Neurology*, 55, 1621–1626.

Duncan, P. W., Weiner, D. K., Chandler, J. and Studenski, S. (1990). Functional reach: a new clinical measure of balance. *Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, 45, M192–M197.

Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.

Hill, K. D., Bernhardt, J., McGann, A. M., Maltese, D. and Berkovits, D. (1996). A new test of dynamic standing balance for stroke patients: reliability, validity, and comparison with healthy elderly. *Physiotherapy Canada*, 48, 257–262.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, 30, 1–15.

Huxham, F. E., Goldie, P. A. and Patla, A. E. (2001). Theoretical considerations in balance assessment. *Australian Journal of Physiotherapy*, 47, 89–100.

Leandri, M., Cammisuli, S., Cammarata, S., Baratto, L., Campbell, M. S. and Tabaton, M. (2009). Balance features in Alzheimer's disease and amnestic mild cognitive impairment. *Journal of Alzheimer's Disease*, 16, 113–120.

Lorbach, E. R., Webster, K. E., Menz, H. B., Wittwer, J. E. and Merory, J. R. (2007). Physiological falls risk assessment in older people with Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, 24, 260–265.

Manckoundia, P., Pfitzenmeyer, P., d'Athis, P., Dubost, V. and Mourey, F. (2006). Impact of cognitive task on the posture of elderly subjects with Alzheimer's disease compared to healthy elderly subjects. *Movement Disorders*, 21, 236–241.

Menant, J. C., Steele, J. R., Menz, H. B., Munro, B. J. and Lord, S. R. (2008). Optimizing footwear for older people at risk of falls. *Journal of Rehabilitation Research and Development*, 45, 1167–1182.

Mitnitski, A. B., Graham, J. E., Mogilner, A. J. and Rockwood, K. (1999). The rate of decline in function in Alzheimer's disease and other dementias. *Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, 54, M65–M69.

NeuroCom^{TM} International Inc. (2003). *Instructions for Use: Balance Master System Operator's Manual, Version 8.1*. Clackamas, OR: NeuroCom International.

Phillips, C. D., Chu, C. W., Morris, J. N. and Hawes, C. (1993). Effects of cognitive impairment on the reliability of geriatric assessments in nursing homes. *Journal of the American Geriatrics Society*, 41, 136–142.

Podsiadlo, D. and Richardson, S. (1991). The Timed Up& Go: a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*, 39, 142–148.

Rolland, Y. *et al.* (2009). An abnormal "one-leg balance" test predicts cognitive decline during Alzheimer's disease. *Journal of Alzheimer's Disease*, 16, 525–531.

Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing reliability. *Psychological Bulletin*, 86, 420–428.

Shumway-Cook, A., Brauer, S. and Woollacott, M. (2000). Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. *Physical Therapy*, 80, 896–903.

Standford, P. W. (2004). Getting more from the literature: estimating the standard error of measurement from reliability studies. *Physiotherapy Canada*, 56, 27–30.

Stratford, P. (1989). Reliability: consistency or differentiating among subjects? *Physical Therapy*, 69, 299–300.

Stratford, P. W. and Goldsmith, C. H. (1997). Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Physical Therapy*, 77, 745–750.

Tappen, R. M., Roach, K. E., Buchner, D., Barry, C. and Edelstein, J. (1997). Reliability of physical performance measures in nursing home residents with Alzheimer's disease. *Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, 52A, M52–M55.

Thomas, V. S. and Hageman, P. A. (2002). A preliminary study on the reliability of physical performance measures in older day-care center clients with dementia. *International Psychogeriatrics*, 14, 17–23.

van Iersel, M., Benraad, C. E. M. and Rikkert, M. G. M. O. (2007). Validity and reliability of quantitative gait analysis in geriatric patients with and without dementia. *Journal of the American Geriatrics Society*, 55, 632–634.

Walter, S. D., Eliasziw, M. and Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine*, 17, 101–110.

Whitney, S. L., Wrisley, D. M., Marchetti, G. F., Gee, M. A., Redfern, M. S. and Furman, J. M. (2005). Clinical measurement of sit-to-stand performance in people with balance disorders: validity of data for the five-times-sit-to-stand test. *Physical Therapy*, 85, 1034.

Wittwer, J. E., Webster, K. E., Andrews, P. T. and Menz, H. B. (2008). Test-retest reliability of spatial and temporal gait parameters of people with Alzheimer's disease. *Gait and Posture*, 28, 392–396.