

Old Dominion University ODU Digital Commons

Physical Therapy and Athletic Training Faculty
Publications

Physical Therapy and Athletic Training

2009

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


Julie D. Ries

John L. Echternach
Old Dominion University

Leah Nof

Michelle Gagnon Blodgett

Follow this and additional works at: https://digitalcommons.odu.edu/pt_pubs

 Part of the [Orthopedics Commons](#), and the [Rehabilitation and Therapy Commons](#)

Repository Citation

Ries, Julie D.; Echternach, John L.; Nof, Leah; and Blodgett, Michelle Gagnon, "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" (2009). *Physical Therapy and Athletic Training Faculty Publications*. 43.
https://digitalcommons.odu.edu/pt_pubs/43

Original Publication Citation

Ries, J. D., Echternach, J. L., Nof, L., & Blodgett, M. G. (2009). 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*, 89(6), 569-579.
doi:10.2522/ptj.20080258

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

Julie D. Ries, John L. Echternach, Leah Nof, Michelle Gagnon Blodgett

Background. With the increasing incidence of Alzheimer disease (AD), determining the validity and reliability of outcome measures for people with this disease is necessary.

Objective. The goals of this study were to assess test-retest reliability of data for the Timed “Up & Go” Test (TUG), the Six-Minute Walk Test (6MWT), and gait speed and to calculate minimal detectable change (MDC) scores for each outcome measure. Performance differences between groups with mild to moderate AD and moderately severe to severe AD (as determined by the Functional Assessment Staging [FAST] scale) were studied.

Design. This was a prospective, nonexperimental, descriptive methodological study.

Methods. Background data collected for 51 people with AD included: use of an assistive device, Mini-Mental Status Examination scores, and FAST scale scores. Each participant engaged in 2 test sessions, separated by a 30- to 60-minute rest period, which included 2 TUG trials, 1 6MWT trial, and 2 gait speed trials using a computerized gait assessment system. A specific cuing protocol was followed to achieve optimal performance during test sessions.

Results. Test-retest reliability values for the TUG, the 6MWT, and gait speed were high for all participants together and for the mild to moderate AD and moderately severe to severe AD groups separately (intraclass correlation coefficients $\geq .973$); however, individual variability of performance also was high. Calculated MDC scores at the 90% confidence interval were: TUG=4.09 seconds, 6MWT=33.5 m (110 ft), and gait speed=9.4 cm/s. The 2 groups were significantly different in performance of clinical tests, with the participants who were more cognitively impaired being more physically and functionally impaired.

Limitations. A single researcher for data collection limited sample numbers and prohibited blinding to dementia level.

Conclusions. The TUG, the 6MWT, and gait speed are reliable outcome measures for use with people with AD, recognizing that individual variability of performance is high. Minimal detectable change scores at the 90% confidence interval can be used to assess change in performance over time and the impact of treatment.

J.D. Ries, PT, PhD, GCS, is Assistant Professor, Program in Physical Therapy, Marymount University, 2807 N Glebe Rd, Arlington, VA 22207 (USA). Address all correspondence to Dr Ries at: julie.ries@marymount.edu.

J.L. Echternach, PT, DPT, EdD, ECS, FAPTA, is Professor and Eminent Scholar Emeritus, School of Physical Therapy, Old Dominion University, Norfolk, Virginia, and Adjunct Professor, Department of Physical Therapy, Nova Southeastern University, Fort Lauderdale, Florida.

L. Nof, PT, PhD, is Professor, Department of Physical Therapy, Nova Southeastern University.

M. Gagnon Blodgett, PsyD, is Clinical Assistant Professor, Department of Geriatrics, Nova Southeastern University.

[Ries JD, Echternach JL, Nof L, Gagnon Blodgett M. Test-Retest reliability and minimal detectable change scores for the Timed “Up & Go” Test, the Six-Minute Walk Test, and gait speed in people with Alzheimer disease. *Phys Ther.* 2009;89:569–579.]

© 2009 American Physical Therapy Association

Post a Rapid Response or find The Bottom Line: www.ptjournal.org

Alzheimer disease (AD) is the most common form of dementia in elderly people and affects an estimated 5.2 million individuals in the United States.¹ It is estimated that 13% of people aged 65 years and older are diagnosed with AD, and the incidence and prevalence increase considerably with age.¹ With the aging of the population, physical therapists in geriatrics will be treating an increasing number of people with AD. Given the need to measure outcomes to assess progress or decline in function, specific clinical tools should be tested for reliability and validity with individuals with AD.

There are recent publications supporting the physical and functional benefits of exercise in the management of AD.^{2,3} Identification of appropriate and useful outcome measures for people with AD would enhance the ability to assess the effectiveness of interventions in clinical and research environments. Our current understanding of the psychometric properties of specific clinical tests with this population is limited. Methodological studies assessing the reliability of clinical tools for people with AD or dementia are scarce, but not nonexistent.⁴⁻⁸ Given the extremely limited research available exclusively with people with a diagnosis of AD, information gleaned from research with individuals with other types of dementia was included in

our review of the literature. Mixed results from studies make it difficult to know which outcome measures will best serve physical therapists' needs in monitoring change in performance in individuals with AD. Outcome measures that have been studied for reliability with individuals with AD or dementia include: the Timed "Up & Go" Test (TUG),^{4,5,8,9} the Six-Minute Walk Test (6MWT),^{4,6} and gait speed.^{4,5,10}

Reliability measurements indicate the degree to which scores of a clinical test are free from measurement errors,¹¹ and although conceptually straightforward, the application of this notion can be complex.^{11,12} Reliability can be expressed as relative reliability or as absolute reliability. If a measurement has high relative reliability, this indicates that repeated measurements will reveal consistent positioning or ranking of individuals' scores within a group.¹¹ If a measurement has high absolute reliability, this indicates that, upon repeated measurement, scores show little variability.¹¹ Relative reliability is measured with correlation coefficients. The intraclass correlation coefficient (ICC) evaluates correlation based upon variance estimates from analysis of variance¹³; the more common the variance between sets of measurements, the higher the ICC.¹² The ICC is an appropriate statistic for examining test-retest reliability.¹³ As a general guideline, an ICC above .75 is considered to demonstrate good reliability; for clinical measures, it is suggested that reliability should exceed .90 to ensure reasonable validity.¹³

Excellent test-retest reliability does not necessarily ensure that individuals' repeated performance will be consistent from test to test. Scores may vary, given expected variability of individual performance and measurement error. A measure of absolute variability provides useful infor-

mation to delineate the "expected" changes from "true" changes in performance. Statistically, absolute reliability is determined by the standard error of measurement (SEM), or the standard deviation of the measurement errors,^{11,13} and a clinically useful mechanism for looking at absolute reliability is the minimal detectable change (MDC) score.¹⁴

Recent literature presenting TUG^{8,10} and gait speed¹⁰ data for individuals with dementia highlights the importance of understanding relative versus absolute reliability. Even though test-retest reliability coefficients for clinical tests are high, individual variability and measurement error make it very difficult to identify a "true" change in performance over time. Minimal detectable change scores provide researchers and clinicians with the opportunity to determine whether a change in performance is a meaningful change (ie, beyond expected measurement error and individual variability).

Clinical observation in people with AD reveals increasing variability of performance with increasing levels of dementia. The existing literature supports this observation. Although Thomas and Hageman⁵ found the TUG to have reasonable test-retest reliability in subjects in day care settings who were considered to have mild to moderate dementia (Mini-Mental Status Examination [MMSE] [SD]=16.9 [7.3]), Tappen et al⁴ found the TUG to be impracticable for use in subjects with moderate to severe AD (MMSE=9.3 [6.0]). Miller et al,⁶ in a *post hoc* assessment of performance on the 6MWT (as a component of assessing test-retest reliability of the Senior Fitness Test), found that subjects who were cognitively impaired showed greater variability than subjects who were cognitively intact; they suggested that the 6MWT is not reliable for use with elderly people who are cognitively



Available With
This Article at
www.ptjournal.org

- [The Bottom Line](#) clinical summary
- [The Bottom Line Podcast](#)
- [Audio Abstracts Podcast](#)

This article was published ahead of print on April 23 2009, at www.ptjournal.org.

impaired. The combined findings of these studies⁴⁻⁶ and the previously noted clinical observation suggest that test-retest reliability of physical and functional performance measures with individuals with AD may be influenced by level of dementia.

The purposes of this research were: (1) to determine test-retest reliability of data for the TUG, the 6MWT, and gait speed with individuals with AD; (2) to determine MDC scores for each of the outcome measures; and (3) to identify performance differences between participant groups stratified by level of dementia.

The existing literature guided the choice of outcome measures for the present study. We hypothesized that the test-retest reliability of the clinical tools would decrease with increased level of dementia, such that the measures would be reliable for use with individuals with mild to moderate AD, but not for use with individuals with moderately severe to severe AD. We also hypothesized that, when stratified by level of dementia, the participants who were less cognitively impaired would perform better on the clinical tests compared with the participants who were more cognitively impaired.

Method

Participants and Procedure

This methodological study used a prospective, nonexperimental, descriptive research design. Guardian informed consent was obtained for all participants, with the exception of one participant who signed her own informed consent statement with her family's approval. When possible, assent forms were signed by participants in conjunction with guardian informed consent. Four sites providing care to individuals with AD (2 inpatient programs and 2 day care programs) participated in the study. The administrative contact at each site aided in the recruitment

of participants. Inclusion criteria were: probable diagnosis of AD, medical stability, and ambulation with or without an assistive device or with handheld guiding assistance of one person. Exclusion criteria were: overt neuromuscular or musculoskeletal problems, acute cardiac or pulmonary conditions, and surgery within the previous 6 months.

Background data were collected primarily from the facility chart and included: age, sex, living environment, and use of an assistive device (classified as "none," "use of a cane," or "use of a walker or rolling walker") or handheld guiding assistance for ambulation. Personal information (eg, vocation, avocations, family members' names, likes and dislikes) was collected from the facility record and staff. This information proved useful in establishing rapport with the participants. The primary researcher (J.D.R.) administered the MMSE to all participants. The primary researcher scored the Functional Assessment Staging (FAST) scale¹⁵⁻¹⁸ using a caregiver or staff informant. The FAST scale has been established as a reliable and valid assessment tool for people with AD.¹⁹ The FAST instrument identifies 16 levels of functioning, separated into 7 stages (Tab. 1), and provided the operational definitions for level of AD in this study. The FAST scale was used to stratify the participants into 2 groups based on level of dementia: a mild to moderate AD group (FAST scale score=4 or 5) and a moderately severe to severe AD group (FAST scale score=6 or 7).

Practical tips on interaction and communication with individuals with AD have been reported in the literature. Every effort was made to integrate these concepts into the protocol for the present study to maximize success of interactions, including: creating a personal connection with the patient using personal

historical information; creating a low-stress environment; using friendly facial expressions, eye contact, and a pleasant, but firm, voice; one-step commands; and stating meaningful goals.²⁰⁻²⁹ The progression of cuing for all tests also was specific and based on the published literature.^{30,31} Cuing began with verbal instruction with a concurrent visual cue or gesture, followed by modeling or demonstration, followed by tactile guidance, and finally, if necessary, physical assistance. Participants were given 10 seconds to respond to a cue before the tester defaulted to a higher level of cuing. A 7-level scale of assistance developed by Beck et al³¹ for elderly people with cognitive impairments was used to classify the type of cuing or assistance required for each participant. If a participant required handheld guiding assistance, every effort was made to allow the participant to drive the movement; however, if the participant stopped or veered from the intended path of movement, the researcher guided the participant's motion back on task.

Two testing sessions for each participant were performed on the same day with a 30- to 60-minute rest period separating testing sessions. Every effort was made to keep all factors associated with the testing sessions consistent (eg, general time of day, staff member assisting with testing, room or area in which testing was performed). Participants performed the TUG, the 6MWT, and the test of gait speed.

The TUG³² is a test of the time required for an individual to stand up from a chair with armrests, walk 3 m, turn, walk back to the chair, and sit down. In the present study, participants circled a small orange cone placed at the 3-m mark. Participants were instructed to "go as fast as you safely can." The stopwatch timing

Table 1.

Functional Assessment Staging (FAST) Scale for People With Alzheimer Disease (AD)^{15,18}

Stage	Level of Functioning	Clinical Diagnosis
1	No decrement	“Normal” adult
2	Subjective deficit in word finding	Normal-aged adult
3	Deficit in demanding employment settings	Compatible with incipient AD
4	Assistance required in complex tasks (eg, handling finances, marketing, planning dinner for guests)	Mild AD
5	Assistance required in choosing proper clothing	Moderate AD
6a	Assistance required in putting on clothing	Moderately severe AD
6b	Assistance required in bathing properly	
6c	Assistance required with the mechanics of toileting (eg, flushing, wiping)	
6d	Urinary incontinence	
6e	Fecal incontinence	
7a	Speech ability limited to approximately a half-dozen intelligible words	Severe AD
7b	Intelligible vocabulary limited to a single word	
7c	Ambulatory ability lost	
7d	Ability to sit up lost	
7e	Ability to smile lost	
7f	Ability to hold up head lost	

started when the participant’s bottom left the chair and ended when the bottom made contact with the chair after the walk.

The 6MWT is the distance walked in a period of 6 minutes. This test was initially considered an endurance measure³³ but more recently has been considered a broader measure of mobility and function.^{34,35} The 6MWT was performed in long hallways of the participating facilities. Participants walked at a “comfortable pace,” were discouraged from talking during the test, and were notified of each passing minute. If participants were distracted or stopped walking, they were prompted to “keep walking” and were advised of the time remaining.

Self-selected gait speed was assessed using the GAITRite walkway.* This portable mat with embedded sensors and companion software creates a

* CIR Systems Inc, 60 Garlor Dr, Havertown, PA 19083.

profile of temporal and spatial parameters of gait and is considered a valid and reliable quantitative gait assessment tool.^{36,37} Participants were instructed to walk at a “comfortable” pace for the length of the mat (4.57 m [15 ft]), and the walking path was established such that acceleration and deceleration did not occur on the mat.

Testing took place at the participating facilities. Patients performed one practice run of the TUG and one practice pass on the GAITRite walkway. They did not perform a practice run of the 6MWT, but were oriented to the walking course. Each testing session included 2 trials of the TUG, 2 passes at a comfortable pace on the GAITRite mat, and 1 trial of the 6MWT. Tests were performed in variable order to control for variability of performance from first to last test as a confounding factor. The order of test administration was randomized, determined by blind drawing of test order from all possible

combinations for each participant. The test order administration remained constant from test session 1 to test session 2 for each participant.

Data Management and Analysis

We used SPSS 15.0 for Windows[†] for data management and analysis. Level of significance was predetermined to be $P < .05$ for all statistical analyses. Descriptive statistics for comparisons of groups included independent-samples *t* tests for parametric data and chi square and Mann-Whitney *U* tests for nonparametric data. All descriptive comparisons between groups were 2-tailed, as no assumptions of directionality were made.

Test-retest reliability of data for all tests was assessed using the ICC (model 2), which is appropriate for methodological research.^{11,13} Reliability of data obtained for the TUG and gait speed was assessed using the ICC (2,2), as mean scores from 2

[†] SPSS Inc, 233 S Wacker Dr, Chicago, IL 60606-6412.

trials from each test session were used in the calculations. Mean scores are considered better estimates of true scores and can increase reliability estimates.¹³ For calculation of test-retest reliability for the 6MWT, the ICC (2,1) was used, as there was only one test score from each session. For each clinical test, the reliability coefficient was calculated for the entire sample and then separately for the mild to moderate AD group and the moderately severe to severe AD group.

Independent-samples *t* tests were used to determine differences in performance on the clinical tests between the 2 groups. Comparisons were made using the mean score of all trials for each participant on the given test (ie, mean of 4 TUG scores, mean of 2 6MWT scores, and mean of 4 gait speed measurements). One-tailed tests were used to assess these data, as there is evidence to suggest that a decrease in speed occurs in patients with dementia³⁸⁻⁴⁵; therefore, an assumption of directionality was thought to be reasonable.

Standard errors of measurement and MDC scores were calculated for the TUG, the 6MWT, and gait speed. Standard errors of measurement¹¹ were calculated using the following equation:

$$(1) \quad SEM = sd \times \sqrt{(1 - r)}$$

In this equation, *sd* is the standard deviation of the measure, and *r* is the reliability coefficient (test-retest reliability in the form of ICC for the subject group). For repeated measures, the SEM was multiplied by the square root of the number of measurements.¹¹

Minimal detectable change scores were calculated for the TUG, 6MWT, and gait speed data at the 90% confidence interval. The formula used for calculating MDC_{90} ^{14,46} was:

$$(2) \quad MDC_{90} = SEM \times 1.65 \times \sqrt{2}$$

In this equation, SEM was calculated as described previously. The 1.65 in the MDC_{90} equation represents the z-score at the 90% confidence level. The product of SEM multiplied by 1.65 is multiplied by the square root of 2 to account for errors associated with repeated measurements.

Results

Data were collected on a total of 53 participants. Two participants' data were eliminated from analysis because their dementia was later determined to be caused by factors other than AD. One setting was not conducive to the performance of the 6MWT, so that test was not performed with participants in that setting. On 2 occasions, individuals at other settings declined to perform the 6MWT. The Figure diagrams a flowchart of participants, explaining any differences between numbers of participants tested and data used in analysis of the results. The remaining 51 participants were stratified into the mild to moderate AD group (*n*=20) and the moderately severe to severe AD group (*n*=31).

Descriptive statistics for the 51 subjects are shown in Table 2. The 2 groups were similar in age, as determined by the independent-samples *t* test, and similar in sex and living environment (ie, home versus inpatient), as determined by analysis of frequencies using the chi-square test. Mini-Mental Status Examination scores also are presented in Table 2. The MMSE scores of the 2 groups were compared using the Mann-Whitney *U* test, a nonparametric statistical analysis, given the ordinal nature of the MMSE data. Given the desire to compare the participants in the present study with those in many published studies that reported MMSE findings using parametric statistics, mean scores and standard deviations are presented for both

groups as well. All participants were consistent in their level of cuing needs from test session 1 to test session 2. The 2 groups were significantly different in their level of cuing needs, as evidenced by Mann-Whitney *U* test analysis, with the participants who were more cognitively impaired requiring higher levels of cuing than those who were less cognitively impaired (Tab. 2). The use of assistive devices (classified as "none," "use of a cane," or "use of a walker or rolling walker") was similar between groups. More than 50% of the participants of both groups were ambulatory without assistive devices. Six participants, all in the moderately severe to severe AD group, required handheld guiding assistance for ambulation.

Intraclass correlation coefficients for test-retest reliability were very high for all outcome measures and for the entire sample and each group (for the TUG, ICC=.985-.988, *P*<.001; for the 6MWT, ICC=.982-.987, *P*<.001; and for gait speed, ICC=.973-.977, *P*<.001).

There were statistically significant differences between the mild to moderate AD group and the moderately severe to severe AD group on TUG, 6MWT, and gait speed performance. The participants who were more cognitively impaired were slower on the TUG and the test of gait speed and walked shorter distances in the 6MWT compared with the participants who were less cognitively impaired (Tab. 3).

Repeated-measures SEMs were calculated for the TUG, the 6MWT, and gait speed to provide a comparison of individual variability of performance across groups (Tab. 4). Although there was little difference in SEMs for the mild to moderate AD group compared with the moderately severe to severe AD group for the 6MWT and gait speed (~10% dif-

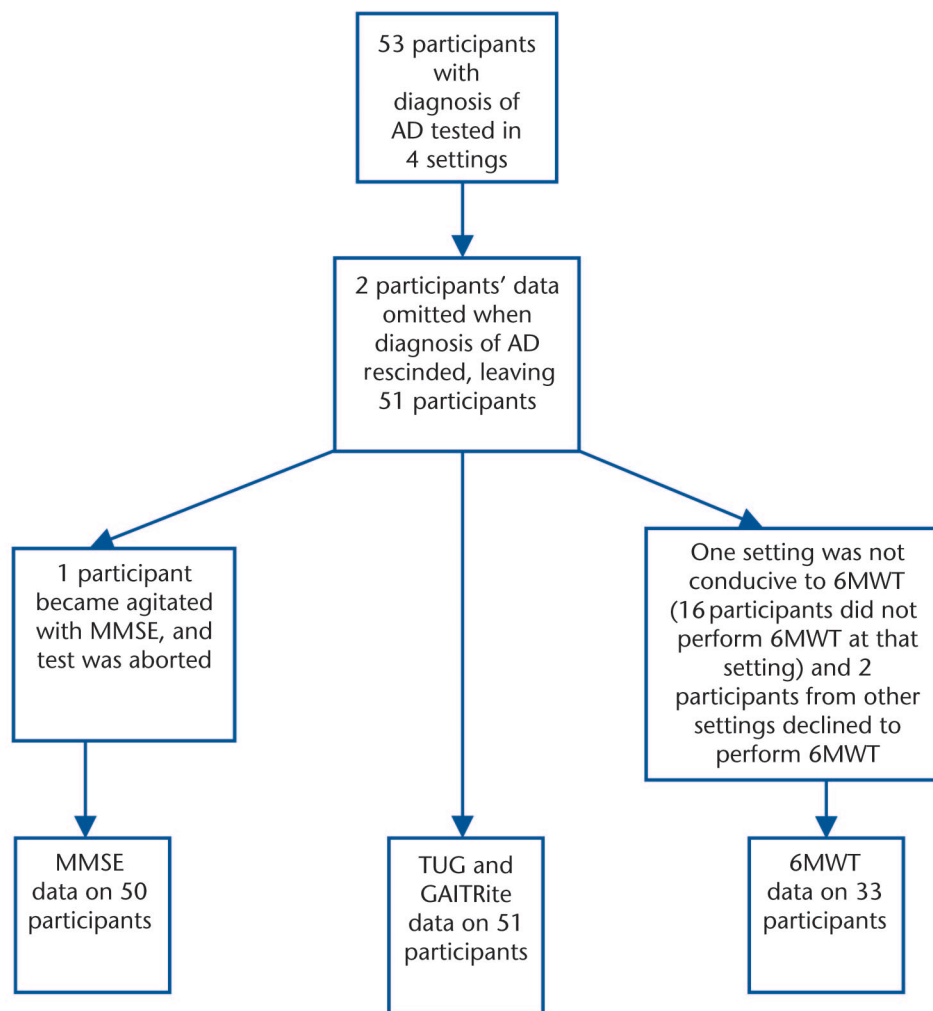


Figure.

Flowchart of study participants. AD=Alzheimer disease, MMSE=Mini-Mental Status Examination, 6MWT=Six-Minute Walk Test, TUG=Timed “Up & Go” Test, GAITRite=computerized walkway test of gait parameters.

ference), there was a substantial difference in SEMs for TUG scores between the 2 groups (~100% difference), with the participants who were more cognitively impaired showing greater variability of performance compared with the participants who were less cognitively impaired. Table 4 also presents the MDC₉₀ values for the TUG, the 6MWT, and gait speed for all participants.

Discussion

Our initial purpose was to determine test-retest reliability of data for the

TUG, the 6MWT, and gait speed for individuals with AD. All 3 outcome measures were found to have excellent test-retest reliability, well exceeding the $r=.90$ threshold¹³ for minimal acceptable reliability for a clinical test and indicating that these tests can be used clinically with good confidence in their test-retest reliability (ie, relative reliability). Test-retest reliability was not influenced by level of dementia, as was hypothesized. Existing literature shows mixed results for relative reliability of these tools in people with AD and dementia.

The TUG appears to be the most widely studied of the tools. We calculated an ICC of .987 for test-retest reliability of TUG scores for all participants. Tappen et al⁴ had such difficulty getting their subjects with moderately severe to severe AD (MMSE=9.3 [6.0]) to perform the TUG, that they had to modify the test beyond recognition. Our participants in the moderately severe to severe AD group had comparable MMSE scores (10.2 [8.8]) and were able to perform the test with excellent relative reliability results. Rockwood et al⁷ reported an ICC of .56

Table 2.
Descriptive Statistics for Participants (N=51)^a

Variable	All Participants (N=51)	Mild to Moderate AD Group (n=20)	Moderately Severe to Severe AD Group (n=31)	Statistical Comparison Between Groups
Age (y), $\bar{X} \pm SD$	80.71 \pm 8.77	81.05 \pm 9.48	80.48 \pm 8.43	$t = .223$ $df = 49$ $P = .824$
Sex (no. female [%])	34 (66.7)	12 (60.0)	22 (70.7)	$\chi^2 = .658$ $df = 1$ $P = .417$
Living environment (no. living at home [%])	39 (76.5)	16 (80.0)	23 (74.2)	$\chi^2 = .228$ $df = 1$ $P = .633$
MMSE $\bar{X} \pm SD$ Range Median	13.1 \pm 8.2 0–30 14	17.40 \pm 4.50 10–26 16.5 Mean rank: 33.78 Sum of ranks: 675.50	10.20 \pm 8.83 0–30 7.5 Mean rank: 19.98 Sum of ranks: 599.50	M-W $U = 134.5$ $P = .001^*$
Levels of assistance for elderly people with cognitive impairment (level of cuing)		Mean rank: 19.05 Sum of ranks: 381.00	Mean rank: 30.48 Sum of ranks: 945.00	M-W $U = 171.00$ $P = .004^*$

^a AD=Alzheimer disease, MMSE=Mini-Mental Status Examination, t =independent-samples t test, df =degrees of freedom, χ^2 =chi-square test of independence, M-W U =Mann-Whitney U test of independent samples. Asterisk indicates statistically significant difference between dementia groups.

for test-retest reliability of TUG scores in elderly individuals with cognitive impairment, but their methodology was fraught with difficulties of working within the limitations of retrospective data. A study by Thomas and Hageman⁵ with a small sample of individuals with mild to moderate dementia (MMSE=16.9 [7.3]) revealed an ICC of .87 for test-retest reliability of TUG scores, and van Iersel et al¹⁰ examined test-retest reliability in people with dementia (MMSE=19.1 [5.2]) and found an ICC of .97 for the TUG. These findings were more consistent

with the higher relative reliability found in our study.

The 6MWT has not been widely used in people with AD or dementia; however, Tappen et al⁴ reported that their participants with AD who were unable to perform the TUG were able to perform the 6MWT. They did not report test-retest reliability of the 6MWT scores, although the research design was such that their ICCs of .76 to .90 for intrarater reliability (one rater observing 2 different sessions) could potentially be interpreted as test-retest reliability. The

authors suggested that the 6MWT may be the preferred test of physical performance for people with AD. We calculated an ICC of .987 for test-retest reliability of 6MWT scores in our study. Thomas and Hageman⁵ and van Iersel et al¹⁰ reported ICCs of .92 and .77, respectively, for test-retest reliability of measurements of self-selected gait speed in people with dementia. We calculated an ICC of .977 for test-retest reliability of gait speed measurements. Our findings consistently showed higher test-retest reliability on all 3 outcome

Table 3.
Performance Differences on Timed “Up & Go” Test, Six-Minute Walk Test, and Gait Speed Between Dementia Groups

Outcome Measure	Dementia Group (n)	$\bar{X} \pm SD$	Independent-Samples t Test, t (df)	P
Timed “Up & Go” Test (s)	Mild to moderate AD (20)	19.95 \pm 9.81	-1.876 (49)	.0335*
	Moderately severe to severe AD (31)	28.01 \pm 17.49		
Six-Minute Walk Test (ft)	Mild to moderate AD (16)	938.78 \pm 428.62	2.411 (31)	.011*
	Moderately severe to severe AD (17)	615.71 \pm 338.34		
Gait speed (cm/s)	Mild to moderate AD (20)	66.07 \pm 29.63	1.823 (49)	.037*
	Moderately severe to severe AD (31)	52.43 \pm 23.58		

^a AD=Alzheimer disease. Asterisk indicates statistically significant at $P < .05$.

Table 4.

Standard Error of Measurement (SEM) for Repeated Measures and Minimal Detectable Change Scores at the 90% Confidence Interval (MDC₉₀) for the Timed “Up & Go” Test, the Six-Minute Walk Test, and Gait Speed

Outcome Measure	SEM			MDC ₉₀ , All Participants
	All Participants (N=51)	Mild to Moderate AD Group (n=20)	Moderately Severe to Severe AD Group (n=31)	
Timed “Up & Go” Test (s)	2.48	1.52	3.03	4.09
Six-Minute Walk Test (m)	20.28 (66.53 ft)	21.86 (71.72 ft)	19.57 (64.20 ft)	33.47 (109.8 ft)
Gait speed (cm/s)	5.72	6.07	5.48	9.44

measures compared with previous research.

One factor that may have enhanced performance on all clinical tests in the present study was the careful use and progression of cuing to facilitate optimal performance. Although we anticipated that performance consistency from one trial to the next perhaps would suffer with increasing dementia, the steady and scripted use of verbal and tactile cuing to optimize performance was carefully implemented; this may have consistently facilitated the best performance. Both Nordin et al⁸ and van Iersel et al¹⁰ commented that the use of cuing was the key to the successful administration of the TUG in subjects with cognitive impairment in their recent reliability studies. In all of the studies reviewed that addressed cuing, the authors either expressed simply a general statement that cuing was allowed^{4,5,10} or reported a dichotomy of cuing versus no cuing.⁸

We believe that careful use of cuing was an asset to consistency of performance, contributing to the high test-retest reliability findings for the clinical tests in our study. Perhaps our careful progression of cuing was pivotal in the successful administration of the TUG in the moderately severe to severe AD group, as Tappen et al⁴ were unable to administer the TUG to their subjects with comparable MMSE scores. Our participants’ cuing needs were rated and

documented (verbal cue/gesture, modeling/demonstration, physical/tactile prompt, progressive amounts of physical guidance) and were consistent from one testing session to the next. Not surprisingly, participants with moderately severe to severe AD required more substantive prompting and guiding for performance of the outcome measures than those with mild to moderate AD. Six of our 51 participants, all from the moderately severe to severe AD group, required handheld guiding assistance of one person to complete the outcome measures. Without the physical guidance of the researcher, these participants would not have been able to complete the tests.

A recent publication by Hauer and Oster⁴⁷ reiterates that measuring functional performance in people with dementia is very complex and cautions researchers that when we provide external cues to patients, perhaps we are measuring the reliability and quality of the external cuing (ie, the researcher’s performance) as opposed to, or as well as, the patients’ performance. In contrast, we contend that a consistent progression of cuing to facilitate best possible performance may be the optimal way to administer clinical tests to people with AD or dementia. The use of a consistent cuing paradigm, in conjunction with following other suggestions related to establishing rapport and maintaining a nonthreatening environment, may allow the

clinician or researcher to repeatedly elicit the most favorable performance from an individual with AD.

Our findings demonstrate that although test-retest reliability (relative reliability) for the clinical tests was excellent, there was still a substantial degree of variability of performance for individual participants from one test session to the next (absolute reliability). The SEM and MDC₉₀ were calculated to objectify these findings. Because the SEM is based on an assumption of normal distribution, probabilities of the normal curve can be applied to SEM values.¹¹ Values from Table 4 can be translated to clinical performance using these principles. For instance, there is a 68% probability that a repeated measure of the TUG will be within ± 1.52 seconds (1 SEM) of the original score for an individual with mild to moderate AD and a 96% probability that a repeated measure will be within 3.04 seconds (2 SEMs) of the original score. For an individual with moderately severe to severe AD, there is a 96% probability that a repeated measure of the TUG will be within ± 6.06 seconds (2 SEMs) of the original score. This could be useful information when examining repeat performances of individuals with AD on the TUG. The dichotomy of dementia levels is important in interpreting clinical findings here, as a difference in performance of approximately 4 to 5 seconds likely represents a change beyond the expected variability in performance in a patient

who is less cognitively impaired, whereas this same change of approximately 4 to 5 seconds would be within the expected variability of performance in a patient who is more profoundly impaired.

The SEM findings for the TUG were consistent with what was anticipated, with the group with a higher level of dementia showing more variability of performance compared with the group with a lower level of dementia. However, this was not the case with the 6MWT or gait speed data. Differences in SEM between groups for the 6MWT and gait speed were small (~10%), with the mild to moderate AD group showing greater variability of performance than the moderately severe to severe AD group. Given the small difference between groups, clinically, it seems appropriate to use the SEM for all individuals if calculating expected performance on repeated measures of the 6MWT and gait speed, irrespective of dementia level. Based on these findings, there is a 96% probability that a repeated measure of the 6MWT will be within ± 40.5 m (133 ft) (2 SEMs) of the initial score. There is a 96% probability that a repeated measure of gait speed will be within ± 11.44 cm/s (2 SEMs) of the initial measurement. These numbers give wide ranges of performance that would fall into the “expected” level of variability on these tests, but still could be clinically useful in the identification of “true” changes in individuals with AD.

The TUG is the only one of the outcome measures we studied that has previously been assessed for absolute reliability. Nordin et al⁸ studied the reliability of TUG scores with participants stratified by cognitive level. As in our study, they hypothesized that increased cognitive impairment would increase the variability of TUG scores, but they found that variability of performance was re-

lated not to cognitive level, but to time to complete the TUG. Also like our study, although their calculated ICCs were high (.91 and .92 for intrarater and intertester reliability, respectively), individual variability also was high. Using logarithmically transformed data, the authors created a calculation for expected variability of TUG performance. This method revealed a large degree of variability or measurement error, such that if an individual performed the TUG in 20 seconds, the expected range of performance on a repeated measure could be between 13.2 and 30.3 seconds. If an individual’s performance was 30 seconds, the expected range of a repeated measure could be between 26.4 and 60.6 seconds. Despite similarities in our general study findings, we used substantially different statistical mechanisms to assess absolute reliability. Our findings suggest that a smaller change in performance on the TUG (ie, 4.09 seconds) than proposed by Nordin et al may represent a clinically significant change. Again, it is possible that our structured and consistent use of cuing and our efforts to maximize participant comfort and minimize environmental stress were effective in minimizing variability of performance, resulting in more consistency across trials.

Our final research goal was to identify performance differences between groups stratified by level of dementia. There were significant differences in performance between the mild to moderate AD group and the moderately severe to severe AD group for the TUG, the 6MWT, and gait speed. The findings of the present study, within the context of published data for the TUG,^{5,10,48,49} the 6MWT,^{4,48,50} and gait speed^{10,51} in individuals with dementia, clearly represent a degradation of performance with the progression of dementia, and this performance decline is beyond that seen with

normal aging. Admittedly, this is piecing together data cross-sectionally; a longitudinal study would be helpful to confirm this observation and would be a useful contribution to the literature.

The present study indicates that the TUG, the 6MWT, and gait speed (using the GAITRite system) are reliable measures for use with individuals with AD. Recently, interpreting change scores and identifying clinically significant changes in performance have become an explicit focus of the physical therapy profession.¹⁴ Clinicians are encouraged to understand how changes in scores translate to clinical relevance. To that end, this study presents MDC_{90} scores that provide meaningful criteria for assessing performance changes for people with AD on the TUG, the 6MWT, and the gait speed test (Tab. 4). Minimal detectable change is the magnitude of change that a measurement must demonstrate to exceed the anticipated measurement error and variability.^{14,46} If a change in score occurs, in either direction, that is greater than MDC_{90} , one can be 90% confident that the difference was not due to measurement error or patient variability. In comparison with the SEM, this provides an even more conservative estimate of a change in score that is clinically meaningful.

Rabheru⁵² recently published a call for the expansion of the mechanism for disease staging and milestones in people with AD, stating that although cognitive milestones are important, functional and behavioral milestones may help to enhance the general picture of the progression of AD. The functional measures in the present study could potentially be a component of the staging process. Van Iersel et al⁴⁴ suggested that a reasonable goal of research should be to identify the minimal clinically important changes in gait variables

in the AD population. The present study suggests that if a gait speed change of greater than 9.44 cm/s is detected in an individual with AD, one can be 90% confident that this represents a “true” change. This information could be helpful in interpreting clinical and research findings related to performance changes in gait speed. This study provides the information to make similar judgments with repeated TUG and 6MWT scores in individuals with AD.

Limitations

There were some limitations of this study. The varied clinical presentation of participants bodes well for the generalizability of the study findings, but the limited geographical (northern Virginia) and socioeconomic (upper middle class) variability of the group may threaten the external validity of the study. The logistics of using a single researcher for data collection influenced sample size and made it impossible to blind the scorer to the dementia level of the patient, which would be ideal in this type of study.

Conclusions

This study demonstrated excellent test-retest reliability for the TUG, the 6MWT, and gait speed in individuals with AD. Significant differences were found in performance of these outcome measures between the mild to moderate AD group and the moderately severe to severe AD group, with the participants who were less cognitively impaired outperforming those who were more cognitively impaired. Despite very high ICCs for test-retest reliability, there was notable individual variability in the performance of these measures. Presentation of SEM and MDC₉₀, for each of the measurement tools, provides clinicians with meaningful thresholds for identifying changes beyond those expected from measurement error and individual variability (ie, “true” change) in individuals with AD. We

are confident that the consistent and scripted use of cuing with participants was instrumental in the successful administration of the outcome measures and contributed to consistency of participant performance across trials. These findings are relevant for both clinicians and researchers. It is our hope that MDC₉₀ values for the TUG, the 6MWT, and gait speed presented here will be helpful in monitoring performance changes over time and assessing the effectiveness of physical therapy and exercise interventions in individuals with AD.

Dr Ries and Dr Echternach provided concept/idea/research design. Dr Ries provided writing, data collection and analysis, project management, participants, facilities/equipment, institutional liaisons, and clerical support. All authors provided consultation (including review of the manuscript before submission).

This research project was approved by the Nova Southeastern University Institutional Review Board (Research Protocol No. HPD-ALL08280604Exp).

This study was conducted in partial fulfillment of the requirements for Dr Ries’ PhD degree in physical therapy from Nova Southeastern University.

A poster presentation of this research was given at the Combined Section Meeting of the American Physical Therapy Association; February 9–12, 2009; Las Vegas, Nevada.

This article was received August 23, 2008, and was accepted March 9, 2009.

DOI: 10.2522/ptj.20080258

References

- 1 2008 Alzheimer’s disease facts and figures. *Alzheimers Dement*. 2008;4:110–133.
- 2 Rolland Y, Pillard F, Klapouszczak A, et al. Exercise program for nursing home residents with Alzheimer’s disease: a 1-year randomized, controlled trial. *J Am Geriatr Soc*. 2007;55:158–165.
- 3 Santana-Sosa E, Barriopedro MI, Lopez-Mojares LM, et al. Exercise training is beneficial for Alzheimer’s patients. *Int J Sports Med*. 2008;29:845–850.
- 4 Tappen RM, Roach KE, Buchner D, et al. Reliability of physical performance measures in nursing home residents with Alzheimer’s disease. *J Gerontol A Biol Sci Med Sci*. 1997;52:M52–M55.

- 5 Thomas VS, Hageman PA. A preliminary study on the reliability of physical performance measures in older day-care center clients with dementia. *Int Psychogeriatr*. 2002;14:17–23.
- 6 Miller E, Whalen R, Biehl J, et al. The reliability of the senior fitness test for assisted living dwellers with different cognitive abilities. Paper presented at: Annual Conference & Exposition of the American Physical Therapy Association; June 30–July 3, 2004; Chicago, Illinois.
- 7 Rockwood K, Awalt E, Carver D, MacKnight C. Feasibility and measurement properties of the Functional Reach and the Timed Up and Go tests in the Canadian Study of Health and Aging. *J Gerontol A Biol Sci Med Sci*. 2000;55:M70–M73.
- 8 Nordin E, Rosendahl E, Lundin-Olsson L. Timed “Up & Go” Test: reliability in older people dependent in activities of daily living—focus on cognitive state. *Phys Ther*. 2006;86:646–655.
- 9 Rockwood K, Strang D, MacKnight C, et al. Interrater reliability of the Clinical Dementia Rating in a multicenter trial. *J Am Geriatr Soc*. 2000;48:558–559.
- 10 van Iersel MB, Benraad CE, Rikkert MG. Validity and reliability of quantitative gait analysis in geriatric patients with and without dementia. *J Am Geriatr Soc*. 2007; 55:632–634.
- 11 Domholdt E. *Rehabilitation Research: Principles and Applications*. 3rd ed. St Louis, MO: Elsevier Saunders; 2005.
- 12 Rothstein JM, Echternach JL. *Primer on Measurement: An Introductory Guide to Measurement Issues*. Alexandria, VA: American Physical Therapy Association; 1993.
- 13 Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 2nd ed. Upper Saddle River, NJ: Prentice Hall Health; 2000.
- 14 Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther*. 2006;86:735–743.
- 15 Reisberg B. Functional assessment staging (FAST). *Psychopharmacol Bull*. 1988;24: 653–659.
- 16 Reisberg B, Ferris SH, Franssen E. An ordinal functional assessment tool for Alzheimer’s-type dementia. *Hosp Community Psychiatry*. 1985;36:593–595.
- 17 Reisberg B, Ferris SH, Anand R, et al. Functional staging of dementia of the Alzheimer type. *Ann NY Acad Sci*. 1984;435: 481–483.
- 18 Reisberg B, Ferris SH, Franssen E. Practical geriatrics: an ordinal functional assessment tool for Alzheimer’s-type dementia. *Hosp Community Psychiatry*. 1985;36: 593–595.
- 19 Sclan SG, Reisberg B. Functional assessment staging (FAST) in Alzheimer’s disease: reliability, validity, and ordinality. *Int Psychogeriatr*. 1992;4(suppl 1):55–69.

- 20 Staples S. Alzheimer's disease: modifying instructions and approach to enhance performance. Paper presented at: Combined Sections Meeting of the American Physical Therapy Association; February 23-27, 2005; New Orleans, Louisiana.
- 21 Haak NJ. Maintaining connections: understanding communication from the perspective of the person with dementia. *Alzheimer's Care Quarterly*. 2002;3:116-131.
- 22 Kovach CR, Henschel H. Planning activities for patients with dementia: a descriptive study of therapeutic activities on special care units. *J Gerontol Nurs*. 1996;22:33-38.
- 23 Patterson JT, Wessel J. Strategies for retraining functional movement in persons with Alzheimer disease: a review. *Physiother Can*. 2002;54:274-280.
- 24 Auer SR, Sclan SG, Yaffee RA, Reisberg B. The neglected half of Alzheimer disease: cognitive and functional concomitants of severe dementia. *J Am Geriatr Soc*. 1994;42:1266-1272.
- 25 Davis CM. The role of the physical and occupational therapist in caring for the victim of Alzheimer's disease. *Phys Occup Ther Geriatr*. 1986;4(3):15-28.
- 26 Hoppes S, Davis LA, Thompson D. Environmental effects on the assessment of people with dementia: a pilot study. *Am J Occup Ther*. 2003;57:396-402.
- 27 Davis LA, Hoppes S, Chesbro SB. Cognitive-communicative and independent living skills assessment in individuals with dementia: a pilot study of environmental impact. *Top Geriatr Rehabil*. 2005;21:136-143.
- 28 Souren LE, Franssen EH, Reisberg B. Neuromotor changes in Alzheimer's disease: implications for patient care. *J Geriatr Psychiatry Neurol*. 1997;10:93-98.
- 29 Oddy R. Promoting mobility in patients with dementia: some suggested strategies for physiotherapists. *Physiother Pract*. 1987;3:18-27.
- 30 Vogelpohl TS, Beck CK, Heacock P, Mercer SO. "I can do it!" Dressing: promoting independence through individualized strategies. *J Gerontol Nurs*. 1996;22(3):39-42; quiz 48.
- 31 Beck CK, Heacock P, Rapp CG, Mercer SO. Assisting cognitively impaired elders with activities of daily living. *American Journal of Alzheimer's Care and Related Disorders & Research*. 1993;8(6):11-20.
- 32 Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc*. 1991;39:142-148.
- 33 Guyatt GH, Sullivan MJ, Thompson PJ, et al. The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure. *Can Med Assoc J*. 1985;132:919-923.
- 34 Harada ND, Chiu V, Stewart AL. Mobility-related function in older adults: assessment with a 6-minute walk test. *Arch Phys Med Rehabil*. 1999;80:837-841.
- 35 Lord SR, Menz HB. Physiologic, psychologic, and health predictors of 6-minute walk performance in older people. *Arch Phys Med Rehabil*. 2002;83:907-911.
- 36 Bilney B, Morris M, Webster K. Concurrent related validity of the GAITRite walkway system for quantification of the spatial and temporal parameters of gait. *Gait Posture*. 2003;17:68-74.
- 37 Menz HB, Latt MD, Tiedemann A, et al. Reliability of the GAITRite walkway system for the quantification of temporospatial parameters of gait in young and older people. *Gait Posture*. 2004;20:20-25.
- 38 Alexander NB, Mollo JM, Giordani B, et al. Maintenance of balance, gait patterns, and obstacle clearance in Alzheimer's disease. *Neurology*. 1995;45:908-914.
- 39 Merory JR, Wittwer JE, Rowe CC, Webster KE. Quantitative gait analysis in patients with dementia with Lewy bodies and Alzheimer's disease. *Gait Posture*. 2007;26:414-419.
- 40 Nakamura T, Meguro K, Yamazaki H, et al. Postural and gait disturbance correlated with decreased frontal cerebral blood flow in Alzheimer disease. *Alzheimer Dis Assoc Disord*. 1997;11:132-139.
- 41 O'Keefe ST, Kazeem H, Philpott RM, et al. Gait disturbance in Alzheimer's disease: a clinical study. *Age Ageing*. 1996;25:313-316.
- 42 Reed R, Yochum K, Alcott S, Meredith K. Gait characteristics of subjects with senile dementia of the Alzheimer's type [abstract]. *J Am Geriatr Soc*. 1992;40:SA79.
- 43 Tanaka A, Okuzumi H, Kobayashi I, et al. Gait disturbance of patients with vascular and Alzheimer-type dementias. *Percept Mot Skills*. 1995;80(3 Pt 1):735-738.
- 44 van Iersel MB, Hoefsloot W, Munneke M, et al. Systematic review of quantitative clinical gait analysis in patients with dementia. *Z Gerontol Geriatr*. 2004;37:27-32.
- 45 Visser H. Gait and balance in senile dementia of Alzheimer's type. *Age Ageing*. 1983;12:296-301.
- 46 Stratford PW. Getting more from the literature: estimating the standard error of measurement from reliability studies. *Physiother Can*. 2004;56:27-30.
- 47 Hauer K, Oster P. Measuring functional performance in persons with dementia. *J Am Geriatr Soc*. 2008;56:949-950.
- 48 Steffen TM, Hacker TA, Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds. *Phys Ther*. 2002;82:128-137.
- 49 Toulotte C, Fabre C, Dangremont B, et al. Effects of physical training on the physical capacity of frail, demented patients with a history of falling: a randomised controlled trial. *Age Ageing*. 2003;32:67-73.
- 50 Roach KE, Tappen RM. Factors related to decline in ambulation ability in institutionalized individuals with Alzheimer's disease [abstract]. *J Geriatr Phys Ther*. 2004;27:119-120.
- 51 Verghese J, Wang C, Lipton RB, et al. Quantitative gait dysfunction and risk of cognitive decline and dementia. *J Neurol Neurosurg Psychiatry*. 2007;78:929-935.
- 52 Rabheru K. Disease staging and milestones. *Can J Neurol Sci*. 2007;34(Suppl 1):S62-S66.