Clinical Utility of the 2-Minute Walk Test for Older Adults Living in Long-Term Care

D.M. Connelly, B.K. Thomas, S.J. Cliffe, W.M. Perry and R.E. Smith

ABSTRACT

Purpose: This study's purposes were to examine the measurement properties of the 2-minute walk test (2MWT), to illustrate the use of reliability coefficients in clinical practice, and to estimate sample size for a subsequent validity study.

Method: Sixteen residents of long-term care (LTC; mean age = 87 years) completed two 2MWTs with Rater A and two 2MWTs with Rater B on test days 1 and 2, approximately 1 week apart. On a third test day, subjects completed one trial of the Berg Balance Scale (BBS), timed up-and-go (TUG) test, and 6-minute walk test (6MWT) with Rater A. On 2 other test days, approximately 1 week apart, Rater A administered the 2MWT to five older adults living in a retirement facility. Absolute and relative reliability and concurrent and known-groups validity coefficients were calculated.

Results: No main effect for rater, trial, or occasion was found. Test-retest reliability estimates of 0.94 and 0.95 were obtained. The 2MWT demonstrated concurrent validity ($r \geq 0.84$) with the BBS, TUG, and 6MWT. Comparison of distance walked by LTC and retirement residents showed a difference of 72.9 m (95% CI: 44.2, 101.6). The results suggest that 90% of truly stable older adults will display random fluctuations in 2MWT performance within a boundary of 15 m.

Conclusion: The 2MWT had sound measurement properties in this sample of LTC residents. Based on our results, 24 subjects would be required for a subsequent hypothesis-testing validity study.

Key Words: 2-minute walk test, frail, older adults, reliability, validity

Connelly DM, Thomas BK, Cliffe SJ, Perry WM, Smith RE. Clinical utility of the 2-minute walk test for older adults living in long-term care. *Physiother Can.* 2009; 61:78-87.

RÉSUMÉ

Objectif : L'étude avait pour objectifs d'évaluer les propriétés de mesure de l'épreuve de marche de 2 minutes, d'illustrer l'emploi de coefficients de fidélité dans la pratique clinique et d'estimer la taille de l'échantillon nécessaire aux fins d'une prochaine étude de validité.

Méthode : Seize résidents d'un centre d'hébergement de soins de longue durée (CHSLD) (moyenne d'âge = 87 ans) ont été soumis à deux reprises à une épreuve de marche de 2 minutes sous la supervision, d'une part, de l'évaluateur A et, d'autre part, de l'évaluateur B aux jours 1 et 2, lesquels ont été espacés d'environ une semaine. À l'occasion d'une troisième journée d'évaluation, les sujets ont été jugés par l'évaluateur A en fonction de leur résultats sur l'échelle d'équilibre de Berg, au test TUG (*timed up-and-go* ou test de la chaise chronométré) et à l'épreuve de marche de 6 minutes. Au cours de 2 autres journées d'évaluation, espacées d'environ 1 semaine, l'évaluateur A a fait subir l'épreuve de marche de 2 minutes à cinq adultes âgés qui résidaient dans une maison pour retraités. Les coefficients de fiabilité relative et absolue, de validité convergente et selon la méthode des groupes connus ont été calculés.

Résultats : Aucun effet notable lié à l'évaluateur, à l'épreuve proprement dite ou au moment où celle-ci a été effectuée n'a été notée. La fiabilité de test-retest a été estimée à 0,94 et à 0,95. La validité convergente de l'épreuve de marche de 2 minutes ($r \geq 0,84$) a été établie avec l'échelle d'équilibre de Berg, le test TUG et l'épreuve de marche de 6 minutes. L'écart quant à la distance de marche entre les résidents du CHSLD et ceux de la maison pour retraités s'élevait à 72,9 m (IC à 95 % : 44,2 – 101,6). Les résultats suggèrent que la distance parcourue à l'épreuve de marche de 2 minutes varie de façon aléatoire de plus ou moins 15 m chez 90 % des personnes âgées vraiment stables.

Conclusion : L'épreuve de marche de 2 minutes possède des propriétés de mesure éprouvées selon cet échantillon de résidents de CHSLD. D'après les résultats, une prochaine étude de validité destinée à vérifier l'hypothèse nécessiterait la participation de 24 sujets.

Mots clés: adultes âgés, épreuve de marche de 2 minutes, fiabilité, fragilité, validité

A version of this paper was presented as an abstract at the World Congress of Physical Therapy in Vancouver, BC, in June 2007: Connelly DM, Thomas B, Cliffe S, Perry W, Smith R. Reliability and validity of the timed two-minute walk test by older adult residents of long-term care centres. *Physiotherapy*. 2007;93(S1):S268.

D.M. Connelly, BScPT, MScPT, PhD: School of Physical Therapy, The University of Western Ontario, London, Ontario.

B.K. Thomas, BScPT, MScPT: School of Rehabilitation Science, McMaster University, Hamilton, Ontario.

S.J. Cliffe, MPT: Out Patient Therapy Services, Joseph Brant Community Health Centre, Burlington, Ontario.

W.M. Perry, MPT: A Body in Motion Rehabilitation, Kitchener, Ontario.

R.E. Smith, MPT: Therapacc Physiotherapy & Rehabilitation, Bobcaygeon, Ontario.

Address for correspondence: Denise M. Connelly, Assistant Professor, School of Physical Therapy, Rm. 1588 Elborn College, 1201 Western Road, The University of Western Ontario, London, ON N6G 1H1;
E-mail: dconnell@uwo.ca.

DOI:10.3138/physio.61.2.78

INTRODUCTION

For older adults, the ability to move from one place to another is one of the most important factors in their perceived levels of health and well-being.¹ Mobility and functional independence are frequent goals of rehabilitation, and, increasingly, pared-down versions of self-report, performance mobility, and function measures are investigated to produce brief but insightful outcome measures. The 2-minute walk test (2MWT), a shorter measure of walking performance than the 6-minute walk test (6MWT),^{2–4} may provide the same information about mobility. The goal of this study is to provide estimates of reliability and validity coefficients for the 2MWT in a sample of older adults.

Previous studies of a variety of adult patient populations reported relative reliability coefficients (intra-class correlation coefficients [ICC]) ranging from 0.82 to 0.99.^{5–10} Reliability, expressed as an ICC, describes the proportion of variance in the scores related to true variance between the objects of measurement (persons, in the current context).¹¹ However, the ICC does not provide information about measurement error in the same units as the original measurement. In contrast, the standard error of measurement (SEM) describes measurement error in the same units as the original measurement and can be applied to provide an estimate of a set of values within which a person's true score is likely to lie.¹² The SEM provides an estimate of absolute reliability and may be helpful to clinicians who use measures to assess change, to compare performance among patients, or to formulate expectations of progression.

Concurrent and known-groups validity of 2MWT performance by adult patients have been described using Pearson r correlations. A range of r values (0.45–0.99) has been reported for 2MWT scores compared to the 10m timed walk and the 4-, 6-, and 12-minute walk tests (MWT).^{10,12–15} When self-report measures of physical function were compared with 2MWT scores, the correlations varied from 0.44 to 0.48.^{7,16} Concurrent validity was reported for 2MWT scores compared to timed up-and-go (TUG) scores ($r = -0.68$ – -0.81)¹⁷ and for the 2MWT versus the 6MWT ($r = 0.95$),¹⁵ but comparisons between the 2MWT and Berg Balance Scale (BBS) scores were not found in the literature. Other types of construct validity reported for 2MWT scores in adults include known-groups,¹⁰ convergent,¹⁵ and discriminant validity.^{15,16} Other validity indexes, including sensitivity, specificity, predictive values, and likelihood ratios, which are useful for interpreting outcome measure results when the measure is used for diagnostic or prognostic purposes,¹⁸ were not found in the literature for the 2MWT.

The 6MWT, TUG, and BBS are commonly used performance measures to evaluate functional status, monitor treatment effectiveness, and estimate prognosis¹⁹

in the older adult population. Together, these three outcome measures represent the underpinnings of independent mobility, defined as the ability to adjust body position with voluntary movement,²⁰ sufficient strength for vertical and horizontal transitions in body position, and aerobic fitness at or above the threshold for community living.²¹ However, these tests have limitations for use with older adults who have multi-system impairments. The 6MWT is often too fatiguing for older adults, and clinicians find it too time-consuming to administer. The TUG is useful for screening purposes but does not require participants to walk distances that would be functional in many settings, such as a long-term care (LTC) facility. Originally developed as an indicator of standing balance, the BBS does not allow individuals to use their regular gait aids and therefore may not accurately reflect their functional day-to-day capabilities. The 2MWT may address some of the limitations of the 6MWT. Therefore, the first purpose of this study was to estimate inter-trial, inter-occasion, and interrater reliability of the 2MWT when administered to older adults. The second purpose was to illustrate how the results from reliability studies can be used to aid clinical decision making. The third purpose was to examine the validity of the 2MWT as an indicator of mobility in older adults and to apply this information to estimate the sample size necessary for a subsequent hypothesis-testing validity study.

METHODS

The study protocol was approved by the facility's Long Term Care (LTC) Committee and by the Research Ethics Board at the University of Western Ontario. All participants provided informed consent.

Subjects

Subjects were recruited from an LTC centre and an affiliated retirement residence, where residents could access occasional nursing care with the assistance of the facility's nursing and recreation staff, respectively. Inclusion criteria for subjects were that they be 70 years of age or older, English speaking, and able to ambulate at least 6 minutes without physical assistance from another person. Exclusion criteria were scores less than 21/30 on the Mini-Mental State Exam (MMSE), uncontrolled heart disease or metabolic disease, resting heart rate ≥ 120 beats per minute, resting blood pressure $> 180/100$, joint replacement within the previous 6 months, hospitalization within the past 30 days, or current involvement in physical rehabilitation. To minimize the effects of possible fatigue or pain, subjects were asked not to participate in vigorous exercise during the 2 hours before their test sessions.

Subjects Living in Long-Term Care

Prior to performance of the physical tests, a chart review was completed for each subject living in the LTC centre to collect health information and demographic data. Included in the chart review form were (1) general demographic data (e.g., age, gender, height, mass); (2) data on number and type of health conditions and medications; and (3) a general health screen to determine whether subjects met any of the exclusion criteria for the study. Subjects also completed a Physical Activity Readiness Questionnaire²² for exercise screening purposes.

Subjects Living in a Retirement Residence

Five retirement-dwelling subjects were interviewed by Rater A to determine their age, height, body mass, number and type of health conditions, and medications, as well as to determine whether they met any of the exclusion criteria for the study.

Raters

The raters were students in the final year of a professional master's degree in physical therapy and had previous experience using the 2MWT, BBS, TUG, and 6MWT. The raters also participated in a session to review and practise using the outcome measures according to standardized procedures.

Study Design

For the reliability component of this study, we applied a repeated-measures design that consisted of trials and occasions. Subjects participated in 3 test days (Table 1). On the first and second test days, LTC subjects completed two trials of the 2MWT with each of two raters. These data were used in the analyses for inter-trial, interrater, and inter-occasion relative and absolute reliability. On the third test day, LTC subjects completed functional measures of mobility, balance, and endurance for comparison with 2MWT results. These measurements were used to assess the concurrent construct validity of the 2MWT. Known-groups construct validity was assessed by comparing distances walked on the 2MWT by the retirement-dwelling and LTC subjects.

Testing

LTC subjects completed testing on 3 days within a 14-day period. All testing occurred at the same time of day, to minimize the influence of meals, fatigue, or leisure and other scheduled activities on subjects' performance. According to standardized verbal instructions, two trials of the 2MWT were conducted independently by each of two raters. The testing order for raters was assigned randomly using a table constructed from coin tosses. A minimum of 5 minutes' rest was provided for subjects between trials, and a minimum of 10 minutes' rest was provided between testing by the two raters. Subjects rested as long as they felt necessary beyond the minimum rest times during and between trials. Raters A and B were blind to each other's measured 2MWT values. Distance walked was measured by raters using a rolling tape measure (Measure Master model MM-12m, Rolatape, Spokane, WA). On each test day, subjects wore the same footwear and used the same gait aid if one was used previously. During a third test day, Rater A administered the BBS, one practice and one test trial of the TUG, and one trial of the 6MWT test, in that order, for each LTC subject. On separate days, testing of retirement-dwelling subjects was completed in the same hallway of the LTC centre. Each retirement-dwelling subject completed two trials of the 2MWT on 2 test days with Rater A, 1 week apart, at the same time of day and wearing the same footwear.

Test Setting

In a quiet, 21 m hallway of the LTC centre, strips of tape were placed on the floor across the width of the hallway to indicate a 3 m distance from a chair with arms for the TUG testing and to mark turnaround points, 1 m from each end of the hallway, for the 2- and 6MWT. Time was measured with a digital hand-held stopwatch (HanHart, Adolf HanHart GmbH & Co KG, Gutenbach, Germany), capable of measuring to the nearest 1/100 of a second. The hallway was brightly lit, with low-pile carpeting on the floor. The hallway was familiar to both LTC and retirement subjects, as it was close to the LTC centre's pool and outdoor patio.

Table 1 Study Design Indicating Raters, Trials, Test Occasions, Outcome Measures, and Origin of Data Used in Reliability and Validity Calculations

	Reliability				Validity		
	Test Day 1 (n = 16 LTC)		Test Day 2 (n = 16 LTC)		Test Day 3 (n = 16 LTC: concurrent validity)	Test Day 4 (n = 5 RR: known-groups validity)	Test Day 5 (n = 5 RR)
	Trial 1	Trial 2	Trial 1	Trial 2			
Rater A	2MWT	2MWT	2MWT	2MWT	BBS, TUG, 6MWT	2MWT	2MWT
Rater B	2MWT	2MWT	2MWT	2MWT			

2MWT = 2-minute walk test; 6MWT = 6-minute walk test; BBS = Berg Balance Scale; LTC = long-term care residents; RR = retirement residents; TUG = timed up-and-go

Outcome Measures

2-Minute Walk Test

The standardized instructions for the 6MWT²³ were modified for use with the 2MWT and explained to subjects at the beginning of the test. Wording was changed to substitute 2MWT for 6MWT. Standardized encouragement was provided to each subject by the rater to ensure that the effect of encouragement on performance was consistent.²⁴ The rater walked behind subjects to ensure their safety and to minimize the effect of pacing during walking.

Subjects started from a clearly marked line of tape on the floor and walked up and down the 21 m hallway from the starting line for 2 minutes at their self-paced walking speed, wearing their regular footwear and using their normal gait aids. Lines on the floor at each end of the hallway indicated where subjects were to turn. Each rater timed two trials using the digital hand-held stopwatch and measured the walked distance using the rolling tape measure, as described above.

Berg Balance Scale

Subjects completed the 14 items of the BBS, according to the guidelines developed by Berg et al.,²⁵ while standing at one end of the 21 m hallway. Raters gave subjects the standardized instructions for each item and provided verbal clarification or demonstration of an item when needed.

Timed Up-and-Go Test

Subjects wore their regular footwear and used their usual gait aids during TUG testing, which was performed at one end of the 21 m hallway. The TUG test began with subjects sitting in an armchair with their backs against the chair back and arms resting on the arms of the chair. Subjects were instructed that on the word “go” they were to stand up, walk at a comfortable and safe pace to a line on the floor 3 m from the chair, turn around, return to the chair and sit down.²⁶ The rater began timing on the word “go” and stopped when subjects were sitting in the chair with their backs against the chair back. Subjects performed one practice TUG prior to the timed trials.

6-Minute Walk Test

Subjects were given the standardized 6MWT instructions and encouragement, as outlined by the American Thoracic Society.²³ Subjects started from a clearly marked line of tape on the floor and walked up and down the 21 m hallway for 6 minutes. Lines on the floor at each end of the hallway indicated where subjects were to turn. Subjects wore their regular footwear and used their normal gait aids. The rater walked behind the subjects to ensure their safety and to minimize the effect of

pacing. Again each test was timed with the digital hand-held stopwatch and the distance was measured using the rolling tape measure.

Subject Safety and Monitoring during Testing

Prior to testing, subjects were fitted with a Polar Vantage NV heart rate (HR) monitor (Polar Electro Oy, Kempele, Finland) around their chest against their skin at the level of the xiphisternum to record instantaneous HR at the onset and cessation of walking during the 2- and 6MWT. The HR was relayed via telemetry to the corresponding wristwatch held by the rater. Subjects rested in a chair for 10 minutes before and after walk tests while a rater measured blood pressure (BP) using a portable automatic BP monitor (Omron model HEM-711ACCAN, Burlington, ON) and recorded resting or recovery HR from the Polar wrist watch.

Subjects were closely monitored as they completed the 2- and 6MWT. For safety, HR during walking could not exceed a sub-maximal rate defined as 75% of age-predicted maximal HR (HR_{max}).²⁷ Before and after the 2- and 6MWT, HR, BP, Rating of Perceived Exertion scores²⁸ for dyspnea and fatigue, and the number and duration of rests taken during the 2- and 6MWT were recorded for participants' safety. The criteria for terminating a walk test included HR above age-predicted 75% HR_{max} target, volitional stopping, or any of the American College of Sports Medicine (ACSM) guidelines for stopping an exercise test.²⁹ As an additional safety precaution, a second student was close by to provide spotting during testing.

DATA ANALYSIS

Relative and Absolute Reliability of 2MWT Performance by LTC Residents

The method described by Eliasziw et al.³⁰ was applied to estimate relative and absolute inter-trial, inter-occasion, and interrater reliability coefficients. Relative reliability (R) was expressed in ICCs. Repeated-measures analysis of variance (ANOVA) was used to estimate the variance components for the reliability coefficients. Depending on the reliability coefficient of interest, the factors were subjects (all analyses), raters (interrater analysis), and trials and occasions (inter-trial and inter-occasion analysis). Because we wished to generalize the reliability beyond the subjects, raters, trials, and occasions in this study, we considered these factors to represent random effects in the ANOVA calculations.¹² Estimates of interrater, inter-trial, and inter-occasion reliabilities were obtained.

Absolute reliability was quantified as the SEM, defined as the standard deviation of the errors in measurement, and was used to interpret an individual's test score.¹² The inter-occasion SEM for Rater A and Rater B was

used to estimate the 90% CI for walking performance (Error_{90}) and the minimal detectable change at a 90% CI (MDC_{90}). The 90% CI for distance walked (Error_{90}) was obtained by multiplying the SEM by 1.65, which is the z -value associated with a 90% CI (i.e., $\text{Error}_{90} = \text{SEM} \times 1.65$). MDC_{90} was obtained by multiplying Error_{90} by the square root of 2 (i.e., $\text{MDC}_{90} = \text{SEM} \times 1.65 \times \sqrt{2}$). The interpretation of MDC_{90} is that 90% of truly stable patients will display random fluctuations less than this value. All data were analyzed using SPSS version 15.0 (SPSS Inc., Chicago, IL). The magnitude of reliability was interpreted according to guidelines offered by Landis and Koch.³¹

Construct Validity of 2MWT Performance by LTC Residents

Concurrent Validity

Pearson r correlations and 95% CIs were calculated to evaluate the associations between 2MWT distance and the other mobility outcome measures. For the group of LTC subjects, individual 2MWT distances measured on the first trial of the first test day were correlated with individual 6MWT distance, TUG test time, and BBS performance values measured during the third test day. The strength of the correlations was characterized according to the guidelines established by Colton.³²

Known-Groups Validity

To determine whether the 2MWT could distinguish between groups of older adults with different levels of walking ability, 2MWT performance was compared between the group of subjects living in a retirement residence and the group living in an LTC centre. Confidence intervals at the 95% level were calculated for the difference in group mean distance walked between the LTC and retirement-dwelling subjects.

Data Analysis for Subjects Who Withdrew from the Study

Group data for the subjects who withdrew from the study were analyzed using descriptive statistics. Independent-sample t -tests were completed to compare data between the subject group and those subjects who withdrew—including group mean age, scores on the MMSE, number of health conditions and medications, and number of falls within the past year. Chi-square statistics were used to investigate whether there were significant differences between the study group and those who withdrew on characteristics including number of men and women in the group, types of gait aids used, and types of health conditions. All comparisons were performed using two-tailed tests, and a difference was considered statistically significant at $p \leq 0.05$.

RESULTS

Subjects

Descriptive analyses were performed to characterize the study group, including group mean age, use of gait aids, MMSE scores, number of health conditions and medications, and number of falls within the past year. Over the duration of the study, 25 subjects from the LTC facility were recruited. Nine subjects from this facility withdrew from the study: three residents refused to complete testing, two had non-symptomatic irregular BP responses during the walk tests and were withdrawn, one did not finish testing because of an outbreak of influenza at the facility, and three were not well enough to complete testing. No significant differences were found between the group of LTC residents who withdrew and those who completed the study in terms of age ($t_{22} = -0.59$, $p = 0.883$), gender ($\chi^2_1 = 4.07$, $p = 0.13$), MMSE scores ($t_{23} = 0.93$, $p = 0.36$), body mass ($t_{18} = 0.82$, $p = 0.42$), or height ($t_{18} = -0.21$, $p = 0.84$).

Complete data were collected for 16 subjects (10 women aged 76–95, mean [SD]: 88.0 [5.4] years, and 6 men aged 76–95, mean: 84.2 [6.6] years; see Table 2) with varied health conditions. The five most common conditions, in descending order of frequency, were cardiovascular disorders, neurological disorders, bone and joint disease, eye disease, and diabetes. The intervals between test days were prolonged for two subjects secondary to quarantine for an influenza outbreak in the LTC centre. For one subject, the first 2 test days were 14 rather than 7 days apart, and for the second subject, the 3 test days were completed in a 17-day rather than a 14-day period. Although neither subject was ill with influenza, both were restricted to their floor of the LTC centre during the outbreak and therefore could not be tested. Visual inspection of the BBS scores and 2- and 6MWT distances for these two subjects determined that their scores were within the group range and were not outliers. No adverse events associated with the mobility and balance testing occurred in the study.

Relative and Absolute Reliability

Table 3 displays descriptive statistics of the 2MWT for trials, test days, and raters. Bland-Altman plots were constructed to provide a visual representation of 2MWT performance between test days³³ as measured by Raters A and B (see Figure 1). The average distance walked on the two trials by individual LTC participants was plotted against the difference score between distance walked on test day 1 and distance walked on test day 2. Table 4 reports the relative and absolute reliability coefficients using data from the first trial.

The SEM for the average of two 2MWT trials on a single test day measured by Rater A was 4.8 m, compared to 4.6 m for the average of first-trial data from test days 1

Table 2 Individual Subject Characteristics for a Group of Older Adults Living in a Long-Term Care Centre ($n=16$) and Another Group of Retirement-Dwelling Older Adults ($n=5$)

Subjects	Age (years)	Height (cm)	Body Mass (kg)	Type of Gait Aid	No. of Prescribed Medications	No. of Chronic Health Conditions	No. of Falls in Past Year	Mini-Mental State Exam (/30)
Long-term care ($n=16$)								
1	95	170.20	53.5	cane	0	1	7	26
2	95	160.00	54.1	rollator	5	5	1	21
3	76	172.50	90.0	rollator	9	6	1	25
4	92	160.00	77.9	rollator	4	7	1	26
5	88	157.50	44.5	rollator	6	4	0	23
6	84	172.70	64.4	rollator	7	4	6	25
7	87	178.70	87.8	rollator	3	5	3	30
8	84	177.00	57.9	rollator	5	4	1	16
9	76	165.10	59.1	rollator	7	2	1	28
10	89	165.00	55.9	rollator	5	5	0	23
11	79	156.90	72.8	rollator	7	5	0	21
12	82	152.40	63.2	none	3	3	3	28
13	88	160.00	53.3	rollator	13	8	3	21
14	89	160.00	65.6	rollator	12	7	4	23
15	89	157	48.2	none	1	1	0	27
16	92	142.20	46.5	rollator	1	1	0	24
Mean (SD), min-max	87 (6.0), 76-95	163.0 (9.6), 142.2-178.7	62.2 (13.8), 44.4-90	13 rollator 1 cane 2 none	6 (4), 0-13	4 (2), 1-8	2 (2), 0-7	24 (4), 16-30
Retirement-dwelling ($n=5$)								
1	83	154.94	48.50	None	5	2	0	23
2	90	165.10	56.70	None	4	4	0	28
3	84	147.32	45.36	None	3	1	0	28
4	81	169.55	60.78	None	1	1	0	29
5	89	149.86	56.25	None	0	0	0	26
Mean (SD), min-max	85 (4.0), 81-90	157.4 (9.6), 147.3-169.6	53.5 (6.4), 45.4-60.8	5 none	3 (2), 0-5	2 (2), 0-4	0	21 (12), 23-29

Table 3 Group Mean (SD) for TUG, BBS, and Distance Walked for Trials, Occasion, and Raters ($n=16$)

	Occasion 1		Occasion 2		Occasion 3	
	Trial 1 (m)	Trial 2 (m)	Trial 1 (m)	Trial 2 (m)	TUG (s)	BBS (/56)
Rater A	77.4 (25.6)	78.4 (23.3)	78.1 (24.5)	77.6 (25.3)	21.0 (9.1)	37 (12)
Rater B	78.5 (23.3)	80.1 (24.3)	80.2 (24.8)	80.0 (23.9)		
Min, max	27.0, 113.1		36.8, 118.6		12.0, 41.2	5, 49
	Rater A	Rater B	Rater A	Rater B		
First trial	77.4 (25.6)	78.5 (23.3)	78.1 (24.5)	80.2 (24.8)

Figure 1 Bland-Altman plots of 2-minute walk distance for trial 1 on test occasion 1 and test occasion 2 for Rater A (A) and Rater B (B). Upper and lower 95% limits of agreement are shown.

Table 4 Relative and Absolute Reliability Coefficients ($n=16$)

	<i>Inter-trial ICC</i>	<i>SEM</i>	<i>Inter-occasion ICC</i>	<i>SEM</i>	<i>Error₉₀</i>	<i>MDC₉₀</i>
Rater A	0.94 (0.87)*	5.8	0.94 (0.88)*	6.3	67.5, 88.3	14.7
Rater B	0.96 (0.91)*	4.6	0.95 (0.91)*	5.2	71.1, 88.3	12.2
First Trial	<i>Inter-rater</i> 0.96 (0.90)*	5.0	<i>Inter-occasion</i> 0.94 (0.89)*	6.0	67.9, 87.7	14.0

*Lower 95% confidence limit

ICC = intraclass correlation coefficient; SEM = standard error of measurement; Error₉₀ = 90% confidence interval for SEM value; MDC₉₀ = minimal detectable change at a 90% confidence level

and 2. SEM values calculated for Rater B's data were 4.4 m for the average of two trials on a single test day and 4.4 m for the average of first-trial data from test days 1 and 2. The SEM was minimally reduced by averaging first-trial data across test days for Rater A, because the subject-by-occasion variance was greater than the subject-by-trial variance. These findings suggest that, from a practical perspective, averaging two trials on a single test day is as good as averaging a single first trial from test days 1 and 2. If the average of two trials on each of 2 test days were used, the SEM would be 3.4 m and 3.1 m for Rater A and Rater B respectively.

The inter-occasion MDC₉₀ values ranged from 12.2 to 14.7 m. The MDC₉₀ is the smallest amount of change in walking performance that can be considered above the threshold of error (SEM) expected in the measurement of an individual's performance. Clinically, this difference between performances would be interpreted as true change and not random measurement error.

Construct Validity

Concurrent Validity

Concurrent validity coefficients for the first trial of the 2MWT are shown in Figure 2. Confidence intervals for the Pearson r values were as follows: TUG (95% CI: -0.59, -0.94); BBS (95% CI: 0.59, 0.94); and 6MWT (95% CI: 0.78, 0.97). Averaged 2MWT distances were similarly correlated to the TUG, BBS, and 6MWT ($r = -0.87, 0.88$, and 0.93 , respectively), as found for the first-trial data of the 2MWT. Because of our relatively small sample, we repeated the correlation analysis applying Spearman's rank correlation coefficient (r_s). This analysis yielded coefficients similar in magnitude to the reported Pearson coefficients.

Known-Groups Validity

Retirement-dwelling older adults walked almost twice as far as residents of LTC on the first trial of 2 test days (retirement group mean [SD]: 150.4 [23.1] m; LTC group mean: 77.5 [25.6] m). The CI for the mean between-group difference of 72.9 m (95% CI: 44.2, 101.6) did not include zero. This provides strong evidence that the walking distances truly differed.

Sample-Size Estimation

Using the example described by Stratford et al.,³⁴ a sample-size estimate was calculated for a future study of the concurrent validity of the 2MWT. The lower CI value from the Pearson r comparison between the 2MWT and the 6MWT was used to set the value for r_{Ha} ; the values used to estimate sample size were $r_{Ha} > 0.78$ and $r_{Ho} = 0.92$ (from Figure 2). The Fisher's z transformation was used to convert the r_{Ho} and r_{Ha} values¹¹ so that they could be used in a sample-size calculation, $N = ((z_\alpha + z_\beta)/\delta)^2 + 3$, where $\delta = |r'_{Ho} - r'_{Ha}|$. The Type I error probability was set at 0.05 one-tailed, and the Type II error probability level was set at 0.20. Based on these assumptions, the sample size for a future hypothesis-testing concurrent validity study was calculated to be 24.

DISCUSSION

Our goals were to estimate the reliability of the 2MWT when administered to older adults and to examine the validity of the 2MWT as an indicator of mobility in older adults. In our sample of older subjects, 2MWT performance measurements were reliable ($ICC \geq 0.94$) and were correlated with subject performances on the BBS, TUG, and 6MWT (Pearson $r \geq 0.84$). Absolute reliability for the 2MWT was calculated as the SEM and estimated to be ≤ 6.3 m. For truly stable subjects, distance walked would randomly fluctuate within a boundary of ± 15 m on subsequent testing. In this study, a test-retest period of 1 week was chosen, in order to minimize the influence of changes in health status during the study period. Also, in order to be confident that performance on the 2MWT was consistent over time, we included only medically stable subjects, minimizing the possibility that change in their health conditions would occur within a 1-week period. Future research on responsiveness to rehabilitation intervention as indicated by the 2MWT should be addressed.

Interpretation of the R and SEM Coefficients

Reliability is a function of the amount of error variance in a set of data³⁵ or the degree to which a measure is free from measurement error.²⁰ Relative reliability,

Figure 2 2-minute walk test distance for trial 1 on test occasion 1 plotted against (a) timed up-and-go scores, (b) Berg Balance Scale total scores, and (c) 6-minute walk test distance measured by Rater A ($n = 16$).

expressed as an ICC, reflects both correlation and agreement between sets of data. Absolute reliability, described as the SEM, is the consistency or stability of repeated responses over time. Response stability is related to measurement error; the SD of the measurement error reflects the reliability of an individual's performance.³⁶ The ICC values in this study (≥ 0.94) are comparable to previous findings for reliability (0.90–0.99) of the 2MWT,⁶ the 6MWT,³⁷ and timed 10-m walk at a comfortable gait speed³⁷ in adults with impaired mobility and older adults living in LTC (0.82–0.89).⁸

The SEM represents inherent variability in stable patients. Miller et al.⁹ reported a MDC_{90} of 19.8 m based on repeated 2MWT performances at a comfortable speed in a group of individuals with impaired mobility secondary to neurological dysfunction. This represents a SEM of 7.1 m, which is slightly higher than the estimates obtained from our sample.

Interpretation of the Pearson r Values

Construct validity values and, in particular, concurrent validity represent the degree to which two measures

that reflect the same construct produce similar results.¹² The correlations between the 2MWT distance and other performance measures of mobility and balance found in this study are similar to previously reported values of validity. In elderly adults with chronic obstructive pulmonary disease, 2MWT distance was found to correlate ($r = 0.95$) with 6MWT distance.¹⁵ Differences in distance walked during the 2MWT have been shown between a control group and a group of adults diagnosed with Parkinson's disease,³⁸ as well as between subjects with stable neurological impairment who did and did not use a gait aid.¹⁰ Comparisons of 2MWT distance and TUG scores have been reported to vary between -0.68 and -0.81 (Pearson r).¹⁷

Study Limitations

A limitation of this study was that subjects represented a sample of convenience. The retirement-dwelling subjects were all women who did not use gait aids, and thus this sample failed to reflect the fact that this segment of the population also includes men and that gait aid use is common. For these reasons, the findings

of this study may not be generalizable to the overall population of older adults living in LTC or retirement residences. Recruitment of participants from one LTC facility, with the inclusion criterion of stable health conditions, limited the number of subjects in the study. Based on the validity data, a future hypothesis-testing study would require 24 subjects.

Clinical Vignette

We present the following vignette to illustrate how the information from the reliability study can be applied to clinical practice. In this example, the patient is an 87-year-old man.

You administer the 2MWT and obtain a value of 80 m. Consider the following questions: (1) How confident are you in the measured value? (2) How much change is required to be reasonably certain that the patient has truly changed?

The answer to the first question is obtained by applying the Error₉₀ estimate. This patient's true walking distance is likely to lie between 69.6 m and 90.4 m (i.e., $80 \pm 1.65 \times \text{SEM}$). The answer to the second question is obtained by referring to the MDC₉₀, which for the 2MWT was 15 m. If this man is one of the 90% of subjects who are truly stable, he will display random fluctuations in walking performance within the bounds of MDC₉₀ (i.e., 65 m and 95 m). Accordingly, a subject who displays a change greater than MDC₉₀ has only a small chance of being a member of the stable subject group and is considered to have truly changed.

Future Research Directions

Because the two subject groups in this study were found to be very different upon comparison of their mobility, a second calculation is required for the purpose of estimating sample size. The comparison groups should be different from each other but more similar than the current comparison to determine whether the 2MWT can detect a smaller difference between two groups. For example, a future comparison might involve community-dwelling users of rolling walkers. These community ambulators may still be different from each other (e.g., higher BBS scores, faster gait speed), but the comparison will be closer because they use the same mobility aid as the majority of our LTC residents. Alternatively, LTC residents who are ambulatory and do not use a gait aid could be compared to retirement-dwelling older adults who do not use a gait aid.

An investigation of the responsiveness of 2MWT distance scores of older adults to a programme of standard inpatient rehabilitation or a community-based strength or aerobic exercise training programme is warranted. In addition, the 2MWT might provide clinically useful information about change in mobility, clinical decision

making, or anticipated recovery of those who have survived a stroke, for example.

CONCLUSION

The 2MWT appears to have sound measurement properties. Physical therapists can obtain an indication of a patient's mobility by averaging two trials of the 2MWT on 1 day or a single trial from 2 test days. The SEM and 90% CI analysis provide clinicians with some guidance for interpretation of patients' performance scores over time.

KEY MESSAGES

What Is Already Known on This Subject

Gait speed and capacity for independent mobility are highly correlated with frailty and sustained aging-in-place.³⁹ Milestones for clinical decision making, expected rehabilitation outcome, and predictors of falls, for example, are integral for clinicians. Evidence-based outcome measures are necessary to mark change in patient performance, can be used to lobby for continued home care service, and may be used to assist with discharge planning for home care or placement to long-term care. Consistency of measurement has been demonstrated for the 2MWT in various patient populations. However, absolute reliability to generate confidence intervals and minimal detectable change has not been investigated.

What This Study Adds

We measured variability of distance walked in a sample of older adult residents of LTC who had multi-system impairments, as represented by their number of prescribed medications, health conditions, accommodation, and use of gait aids. SEM values in metres were generated using single trial, averaged trials, and trials averaged across test days. The results of this study suggest that clinicians can use a single measure of 2MWT distance per visit to monitor patient progression. The 2MWT appears to provide information similar to the 6MWT for this group of older adults.

ACKNOWLEDGMENT

The authors wish to acknowledge S. Hammond, T. Stevenson, T. Overend, D. Bryant, and P. Stratford for their assistance with this study.

REFERENCES

1. Bourret EM, Bernick LG, Cott CA, Kontos PC. The meaning of mobility for residents and staff in long-term care facilities. *J Adv Nurs*. 2002;37:338–45.
2. Enright P. The six-minute walk test. *Respir Care*. 2003;48:783–5.

3. Harada ND, Chiu V, Steward AL. Mobility-related function in older adults: assessment with a 6-minute walk test. *Arch Phys Med Rehabil*. 1999;80:837–41.
4. Steffen TM, Hacker TA, Mollinger LA. 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*. 2003;82:128–37.
5. Bean JF, Kiely DK, Leveille SG, Herman S, Huynh C, Fielding R, *et al*. The 6-minute walk test in mobility-limited elders: what is being measured? *J Gerontol A-Biol*. 2002;57:M751–6.
6. Brooks D, Hunter J, Parsons J, Livsey E, Quirt J, Devlin M. Reliability of the two-minute walk test in individuals with transtibial amputation. *Arch Phys Med Rehabil*. 2002;83:1562–5.
7. Brooks D, Parsons J, Tran D, Jeng B, Gorczyca B, Newton J, *et al*. The two-minute walk test as a measure of functional capacity in cardiac surgery patients. *Arch Phys Med Rehabil*. 2004;85:1525–30.
8. Connelly DM, Stevenson TJ, Vandervoort AA. Between- and within-rater reliability of walking tests in a frail elderly population. *Physiother Can*. 1996;48:47–51.
9. Miller P, Moreland J, Stevenson TJ. Measurement properties of a standardized version of the two-minute walk test for individuals with neurological dysfunction. *Physiother Can*. 2002;54:241–57.
10. Rossier P, Wade DT. Validity and reliability comparison of four mobility measures in patients presenting with neurologic impairment. *Arch Phys Med Rehabil*. 2001;82:9–13.
11. Norman GR, Streiner DL. *Biostatistics: the bare essentials*. 2nd ed. Hamilton, ON: BC Decker; 2000.
12. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. 3rd ed. Upper Saddle River, NJ: Pearson Education; 2007.
13. Kosak M, Smith T. Comparison of the 2-, 6-, and 12-minute walk tests in patients with stroke. *J Rehabil Res Dev*. 2005;42:103–7.
14. Leung AS, Chan KK, Sykes K, Chan KS. Reliability, validity and responsiveness of a 2-min walk test to assess exercise capacity of COPD patients. *Chest*. 2006;130:119–25.
15. Bernstein ML, Despars JA, Singh NP, Avalos K, Stansbury DW, Light RW. Reanalysis of the 12 minute walk in patients with chronic obstructive pulmonary disease. *Chest*. 1994;105:163–7.
16. Brooks D, Parsons J, Hunter JP, Devlin M, Walker J. The 2-minute walk test as a measure of functional improvement in persons with lower limb amputation. *Arch Phys Med Rehabil*. 2001;82:1478–83.
17. Brooks D, Davis AM, Naglie G. Validity of 3 physical performance measures in inpatient geriatric rehabilitation. *Arch Phys Med Rehabil*. 2006;87:105–10.
18. Riddle DL, Stratford PW. Interpreting validity indexes for diagnostic tests: an illustration using the Berg balance test. *Phys Ther*. 1999;79: 939–48.
19. Solway S, Brooks D, Lacasse Y, Thomas S. A qualitative systematic overview of functional walk tests used in the cardiorespiratory domain. *Chest*. 2001;119:256–70.
20. Finch E, Brooks D, Stratford PW, Mayo NE. *Physical rehabilitation outcome measures: a guide to enhanced clinical decision making*. 2nd ed. Hamilton, ON: BC Decker; 2002.
21. Paterson DH, Cunningham DA, Koval JJ, St Croix CM. Aerobic fitness in a population of independently living men and women aged 55–86 years. *Med Sci Sport Exerc*. 1999;31:1813–20.
22. Canadian Society for Exercise Physiology [homepage on the Internet]. Ottawa: The Society; 2002 [updated 2002; cited 2009 Feb 4]. Physical activity readiness questionnaire (PAR-Q) [PDF document]. Available from: <http://www.confmanager.com/main.cfm?cid=574&nid=5110>.
23. American Thoracic Society [ATS]. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med*. 2002;166:111–7.
24. Guyatt GH, Pugsley SO, Sullivan MJ, Thompson PJ, Berman L, Jones NL, *et al*. Effects of encouragement on walking test performance. *Thorax*. 1984;39:818–22.
25. Berg K, Wood-Dauphinee S, Williams JL, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. *Physiother Can*. 1989;41:304–11.
26. Podsiadlo D, Richardson S. The timed “up & go”: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc*. 1991; 39:142–8.
27. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. *J Am Coll Cardiol*. 2001;37:153–6.
28. Borg G. *Borg’s perceived exertion and pain scales*. Champaign, IL: Human Kinetics; 1998.
29. American College of Sports Medicine [ACSM]. *ACSM’s guidelines for exercise testing and prescription*. 6th ed. Philadelphia: Lippincott Williams & Wilkins; 2000.
30. Eilaszew M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Phys Ther*. 1994;74:777–88.
31. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74.
32. Colton T. *Statistics in medicine*. Boston: Little, Brown; 1974.
33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327: 307–10.
34. Stratford PW, Binkley JM, Stratford DM. Development and initial validation of the upper extremity functional index. *Physiother Can*. 2001;53:259–67.
35. Fisher RA. *Statistical methods and scientific inference*. Edinburgh: Oliver & Boyd; 1956.
36. Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. *Designing clinical research*. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2001.
37. Flansbjerg U, Holmba AM, Downham D, Patten C, Lexell J. Reliability of gait performance tests in men and women with hemiparesis after stroke. *J Rehabil Med*. 2005;37:75–82.
38. Light KE, Behrman AL, Thigben M, Triggs WJ. The 2-minute walk test: a tool for evaluating endurance in clients with Parkinson’s disease. *Neurol Rep*. 1997;21:136–9.
39. Montero-Odasso M, Schapira M, Soriano ER, Varela M, Kaplan R, Camera LA, *et al*. Gait velocity as a single predictor of adverse events in healthy seniors aged 75 years and older. *J Gerontol A-Biol*. 2005;60:1304–9.