

Research Article

# Reliability and Minimal Detectable Change Values for Performance-Based Measures of Physical Functioning in the Canadian Longitudinal Study on Aging

Marla K. Beauchamp, PhD,<sup>1,\*</sup>† Qiukui Hao, MD,<sup>1,2,†</sup> Ayse Kuspinar, PhD,<sup>1</sup> Cassandra D'Amore, BSc,<sup>1</sup> Giulia Scime, MSc,<sup>3</sup> Jinhui Ma, PhD,<sup>4</sup> Alexandra Mayhew, PhD,<sup>4</sup> Carol Bassim, MSc, DMD,<sup>4</sup> Christina Wolfson, PhD,<sup>5,6,7</sup> Susan Kirkland, PhD,<sup>8</sup> Lauren Griffith, PhD,<sup>4</sup> and Parminder Raina, PhD<sup>4</sup>

<sup>1</sup>Faculty of Health Sciences, School of Rehabilitation Science, McMaster University, Hamilton, Ontario, Canada. <sup>2</sup>The Center of Gerontology and Geriatrics/National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University, Chengdu, China. <sup>3</sup>Canadian Longitudinal Study on Aging, Hamilton Data Collection Site, Ontario, Canada. <sup>4</sup>Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada. <sup>5</sup>Department of Medicine, Faculty of Medicine and Health Sciences, McGill University, Montreal, Quebec, Canada. <sup>6</sup>Department of Epidemiology, Biostatistics and Occupational Health, School of Population and Global Health, McGill University, Montreal, Quebec, Canada. <sup>7</sup>Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada. <sup>8</sup>Department of Community Health and Epidemiology and Division of Geriatric Medicine, Faculty of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada.

\*Address correspondence to: Marla K. Beauchamp, PhD, School of Rehabilitation Science, Institute of Applied Health Sciences, McMaster University, Hamilton, ON L8S 4K1, Canada. E-mail: [beaucm1@mcmaster.ca](mailto:beaucm1@mcmaster.ca)

†Cofirst authors.

Received: February 24, 2021; Editorial Decision Date: June 7, 2021

**Decision Editor:** Lewis Lipsitz, MD, FGSA

## Abstract

**Background:** The aim of this study was to determine the relative and absolute reliabilities of 5 key performance-based measures of physical function in the Canadian Longitudinal Study on Aging (CLSA).

**Methods:** An age-stratified subsample of 147 participants from the CLSA who were undergoing their 3-year data collection visit participated in 2 repeat visits (within 1 week). Participants underwent tests of grip strength, 4-m gait speed, Timed Up and Go (TUG), chair rise, and single-leg stance (left, right, mean, maximum). Intraclass correlation coefficients (ICCs), standard error of measurement, and minimal detectable change (MDC) values were calculated.

**Results:** The relative reliability for grip strength was excellent (ICC = 0.95); the TUG and single-leg stance tests had good reliability (ICC = 0.80 or 0.78–0.82, respectively); gait speed and the chair-rise test had moderate reliability (ICC = 0.64 for both) for participants overall. For participants between 50 and 64 years, TUG and gait speed had poor reliabilities (ICC = 0.38 or 0.33, respectively). For participants aged 75 years and older, the single-leg stance had poor reliability (ICC = 0.30–0.39). The MDC<sub>90</sub> was about 6 kg for grip strength, 2.3 seconds for TUG, 0.2 m/second for gait speed, 5.2 seconds for chair rise, and ranged from 22.8 to 26.2 seconds for the single-leg stance.

**Conclusions:** Among community-dwelling Canadians older than 50 years, the reliabilities of the CLSA measures were moderate to excellent. The TUG and gait speed in the youngest age group, and the single-leg stance in the oldest age group, showed poor reliability. MDC values can be used to interpret changes over time.

**Keywords:** Balance, CLSA, Mobility, Muscle, Rehabilitation

Performance-based measures (PBMs) of physical function are commonly used to identify older adults at risk for functional decline, disability, and death and for monitoring change over time (1–4). Among the most frequently used PBMs are simple tests of walking, balance, and strength, such as the 4-m gait speed test, Timed Up and Go (TUG) test, chair-rise test, single-leg stance test, and measures of grip strength. All these tests are included as part of the core physical performance battery in the Canadian Longitudinal Study on Aging (CLSA), a large-scale, 20-year study addressing the complexities of aging in over 50 000 participants. Although there are many reports on the psychometric properties of PBMs, there remain knowledge gaps with respect to their reliability and optimal change thresholds—data that are critical for enabling reliable quantification of change over time in studies such as the CLSA.

Much of the prior work on test–retest reliability of commonly used PBMs has been limited to convenience samples involving narrow age groups and/or specific clinical populations (5–10). For example, for isometric tests of grip strength, Bohannon and Schaubert (9) reported an intraclass correlation coefficient (ICC) for test–retest reliability of 0.91 for the left hand and 0.95 for the right hand. However, these results were based on a convenience sample of only 21 community-dwelling older adults (mean age 75 years, range: 65–85 years), and measurements were taken over a 12-week period. Similarly, although the developers of the TUG (8) originally reported an ICC of 0.99 for 20 older adults (mean age 80 years, range: 60–90 years) attending outpatient rehabilitation, Rockwood et al. (10) noted a test–retest reliability of only 0.56 in a larger sample of older adults with and without cognitive impairment ( $n = 2305$ , mean age 78 years, range: 69–104 years), and Jette et al. (7) found an ICC of 0.74 in frail older adults ( $n = 105$ , mean age 78 years, range: 65–94 years) enrolled in a clinical drug trial. For chair rise, gait speed, and single-leg stance, the ICCs reported have ranged between 0.67 and 0.99 in previous studies of community-dwelling people aged 60 years or older (11–16). In addition, although age-related declines in performance on these tests have been noted from as young as age 50 (17), to the best of our knowledge, no previous studies have examined the reliability of PBMs stratified by age, especially included individuals aged 50–64 years. Precise estimates of reliability are also needed to determine thresholds of true and important change for PBMs in community-dwelling adults—an area with limited research to date despite their widespread use.

The aim of this study was to determine the reliability and minimal detectable change (MDC) values of commonly used PBMs in middle-aged and older adults enrolled in the CLSA. The specific objectives are to (a) determine the relative and absolute reliabilities of the TUG, gait speed, grip strength, chair-rise, and single-leg stance tests in an age-stratified sample of community-dwelling adults and (b) estimate MDC thresholds for these measures. These data are critical to be able to adequately characterize changes in physical functioning among older adults in longitudinal analyses of the CLSA data.

## Method

### Study Design and Participants

This was a prospective test–retest reliability substudy of the CLSA (Hamilton Data Collection Site). The CLSA is a population-based, 20-year, prospective cohort study collecting data every 3 years on over 50 000 men and women aged 45–85 years at enrollment (18,19). For the present study, we enrolled an equal number of participants

in 3 age groups: 50–64, 65–74, and 75–90 years. Inclusion criteria were that participants had to be living at home and have the ability to understand simple instructions. We excluded participants with any contraindications to do the performance-based tests (as per the CLSA protocol [<https://www.clsa-elcv.ca/researchers/physical-assessments>]). Participants were consecutively enrolled until the required sample size was met in each age strata. Participants took part in a baseline assessment session for this substudy as part of their routine CLSA visit and were asked to return for repeat testing approximately 1 week later. The study protocol was approved by the Hamilton Integrated Research Ethics Board (2018-5280-GRA).

### Study Process

Consecutive eligible CLSA participants undergoing their in-home face-to-face interviews were approached by a member of the research team to determine their interest in participating in the reliability substudy, from March to October 2019. Those who expressed interest were telephoned prior to attending their CLSA site-based data collection visit to explain the study procedures and obtain verbal consent. At their data collection visit, participants who consented underwent the usual CLSA battery of tests (Time I). Performance on the PBMs was recorded in a separate database for this substudy. In order to fully describe the sample, the following information was extracted from the core CLSA data on each participant: age, sex, weight, height, income, living situation, education, and self-reported physician-diagnosed chronic conditions (eg, hypertension, cardiovascular disease, diabetes, vision disease, musculoskeletal disease, neurological disease, and mental health disease). We also recorded participants' history of falls in the past 12 months. One week later (Time II), participants returned to the Data Collection Site and repeated the physical performance battery with the same order. CLSA research staff or postgraduate students in rehabilitation administered the tests and raters were not standardized between visits. Before data collection, all raters (CLSA research staff or postgraduate students in rehabilitation) received at least one formal training session according to the CLSA standard operating procedures (<https://www.clsa-elcv.ca/researchers/physical-assessments>). Participants were also asked to indicate whether they had perceived any change in their physical health from Time I to Time II by the following question: "Since your last visit, how would you rate your physical health?" with 5 response options: much worse, slightly worse, about the same, slightly better, much better.

### Measures

Five commonly used PBMs: TUG, gait speed, single-leg stance test, grip strength, and chair-rise test were conducted at Time I (CLSA assessment) and Time II (retest) approximately 1 week later. Tests were performed according to the CLSA standard operating procedures (<https://www.clsa-elcv.ca/researchers/physical-assessments>).

Grip strength is a quantitative measure of isometric muscle strength of the hand and forearm. Standardized instructions and positioning were used during testing. Three consecutive trials for the dominant hand were conducted, and the mean and maximum values were recorded (20).

The TUG is a commonly used PBM of functional mobility and balance among older adults (21). Participants were asked to rise from a chair, walk 3 m, turn around, walk back to the chair, and sit down (chair height 46 cm). The test was performed by asking participants to "walk at your normal pace." Participants were permitted to use the assistive device they typically use in the community. The timer was started

immediately after the evaluator said the command “go” and stopped when the participant returned to sitting in the chair.

The chair-rise test, also called the Five-Times-Sit-to-Stand test, measures lower limb strength and balance (22). Participants were asked to cross their arms on their chest. The evaluator provided the following standardized instructions: “I want you to stand up and sit down 5 times as quickly as you can when I say ‘Go.’” Timing begins when the evaluator says “Go” and stops when the participant is fully standing for the fifth time.

The 4-m gait speed test measures walking speed in meters per second (m/s) (11). Participants were asked to walk 4 m at their usual pace, according to the following instructions “after I say, ‘ready, set, go’, please walk at your usual walking pace until I say to stop.” The timer was started immediately after the evaluator said “Ready, set, go” and stopped when the participant was completely across the finish line.

The single-leg stance test measures an individual’s static balance. Participants were instructed to stand on one leg for as long as possible starting with the right leg. The test was then repeated on the left leg. Timing began when the foot left the ground on the first attempt and stopped when the foot touched the ground or when the participant lost balance or touched the wall. Data were recorded from each trial, as were the mean and maximum of the 2 trials.

### Statistical Analysis

To characterize the study participants, number and percentages were used for describing categorical variables; means, standard deviations and/or median, quartiles (Q1, Q3) were used for continuous variables depending on the distribution. We assessed the relative reliability by calculating ICC<sub>2,1</sub> and absolute reliabilities using the standard error of measurement (SEM) and MDC for each of the performance-based tests (23,24). The ICC<sub>2,1</sub> estimates and 95% confidence intervals were calculated using “irr” R package based on a single-rater, absolute-agreement, 2-way random-effects model using the following formula:  $ICC_{2,1} = (MS_R - MS_E) / [MS_R + MS_E + 2(MS_C - MS_E) / n]$ . In the formula,  $MS_R$  is the mean square for rows (between-participant variability),  $MS_E$  is the mean square for error (residual variability),  $MS_C$  is the mean square for columns (between-test variability), and  $n$  is the number of participants. Values for ICC greater than 0.90 indicate excellent reliability, between 0.75 and 0.9 indicate good reliability, between 0.5 and 0.75 indicate moderate reliability, and lower than 0.5 indicate poor reliability (23).

For measures that exhibited poor reliability, we conducted a variance component analysis to estimate the contribution of each considered variable to the variance of the dependent variable. In the current context, the variance in the measurement scores might arise from between-participant variability, between-testing variability, between-raters variability, and residual variability (errors). We assumed that the raters and participants in the sample represent random selections from larger populations. We performed the variance component analysis using “VCA” R package (25) with the analysis of variance method by putting age, participants, different test times (Time I or Time II), and raters as random factors in the models. Moreover, we used Bland–Altman plots with 95% limits to visualize agreement between the 2 repeated measurements and identify extreme values for measures that exhibited poor reliability (26). If extreme values were found in the Bland–Altman plots, we excluded extreme values for participants who reported a change in their physical health between the 2 assessments in order to examine how these extreme values influenced the ICCs.

The SEM was calculated as  $SD\sqrt{1-ICC}$  and the MDC with 90% confidence was calculated as  $SEM \times \sqrt{2} \times 1.65$  in each age group. The interpretation of the  $MDC_{90}$  is that it is the smallest change in a measure that can be considered real change beyond measurement error with 90% confidence (27). Statistical analyses and graph construction were performed using the software R 4.0.2 (R Core Team, 2020) in RStudio (version 1.2.1335).

The sample size for the study was based on an expected ICC value of 0.80 with a 95% CI of  $\pm 0.1$ . Fifty participants for each of the 3 age groups were targeted for a total sample size of 150 participants (28).

## Results

### Study Population and Baseline Characteristics

A total of 151 participants agreed to participate in the substudy. For this analysis, 4 participants had incomplete data and were excluded, resulting in a study sample of 147 (76 males and 71 females). For each PBM, only a few data values were missing at Time I or II (up to 7 cases) due to safety concerns (high risk of fall), so missing data were not imputed in our analysis. The mean age of the participants was 69 (10) years (range 51–90), with 3 age strata: 50–64 years ( $n = 48$ ), 65–74 years ( $n = 50$ ), and 75 and older ( $n = 49$ ).

The median time between the 2 tests was 7 days. Table 1 presents the demographic characteristics of the participants. Most participants had a postsecondary degree; few participants used a gait aid at home and/or in the community. Generally, older participants had a higher prevalence of fall history in the past year and were more likely to report living alone, more chronic conditions, and more medications.

### Relative Reliability of Measures

Test–retest reliability values for grip strength, TUG test, gait speed, chair-rise test, and single-leg stance test for all participants are given in Table 2. Overall, the ICCs for grip strength (mean and maximum) were excellent (ICC = 0.95, 95% CI: 0.92–0.97); good for TUG (ICC = 0.80, 95% CI: 0.72–0.86) and single-leg stance tests (ICC for right, left, mean, and maximum = 0.78–0.82, 95% CI: 0.70–0.87); and moderate for gait speed and chair-rise test (ICC = 0.64, 95% CI for gait speed: 0.54–0.73; 95% CI for chair-rise test: 0.45–0.77).

Test–retest reliability values for the PBMs according to age groups are given in Table 3. Grip strength (mean and maximum) had excellent reliability across all 3 age groups (ICC = 0.91–0.97). TUG and gait speed had poor reliability (ICC = 0.38, 95% CI: 0.12–0.59; 0.33, 95% CI: 0.05–0.55, respectively) among participants aged 50–64 years. Single-leg stance test (right, left, mean, and maximum) had poor reliability among participants aged 75 years and older (ICC = 0.30–0.39, 95% CI: 0.02–0.61). Bland–Altman plots for the measures that had poor reliability in specific age groups are shown in Figure 1. Most of the participants (124/147) reported no change in their physical health between Time I and Time II, 7/48 participants reported a change (better or worse) in the 50–64 age group, 9/50 in 65–74 age group, and 7/49 in 75 and older age group. There are 3 extreme values of TUG in those aged 50–64 years (Figure 1A); 2 of these reported slightly worse physical health at the Time II test compared with their Time I test. After removing the 2 extreme values, we did not find a substantial increase in the ICC (ICC = 0.43, 95% CI: 0.14–0.65). Similarly, this kind of extreme value did not influence our results for the gait speed test. Although there were several extreme values for the single-leg stance test, participants with these

**Table 1.** Demographic Characteristics of the Participants

	50–64 years N = 48	65–74 years N = 50	75+ years N = 49	All Participants N = 147
Age (years), mean (SD)	58 (4)	69 (3)	81 (4)	69 (10)
Female, <i>n</i> (%)	30 (62.5)	21 (42.0)	20 (40.8)	71 (48.3)
Days between 2 tests, median (Q1, Q3)	7 (7, 9)	8 (7, 10)	7 (7, 8)	7 (7, 9)
Height (cm), mean (SD)	170 (9)	170 (9)	167 (10)	169 (10)
Weight (kg), mean (SD)	85 (17)	83 (15)	77 (17)	82 (17)
BMI (kg/m <sup>2</sup> ), mean (SD)	29.1 (4.8)	28.9 (5.3)	27.7 (5.8)	28.6 (5.3)
Use of gait aid, <i>n</i> (%)	0	6 (12.0)	5 (10.2)	11 (7.4)
Balance/fall history, <i>n</i> (%)				
Difficulty with balance	10 (20.8)	12 (24.0)	16 (32.7)	38 (25.9)
Fall in the last year	9 (18.8)	11 (22.0)	12 (24.5)	33 (21.8)
Worry about falling	10 (20.8)	15 (30.0)	20 (40.8)	45 (30.6)
Injury due to fall	6 (12.5)	6 (12.0)	8 (16.3)	20 (13.6)
Number of medications, median (Q1, Q3)	1 (0, 3)	2 (1, 4)	4 (2, 6)	2 (1, 4)
Chronic conditions (number of participants), <i>n</i> (%)				
Hypertension	12 (25.0)	22 (44.0)	27 (55.1)	61 (41.5)
Cardiovascular disease	4 (8.3)	7 (14.0)	25 (51.0)	36 (24.5)
Diabetes	6 (12.5)	10 (20.0)	8 (16.3)	24 (16.3)
Vision disease (eg, cataracts)	6 (12.5)	18 (36.0)	35 (71.4)	59 (40.1)
Musculoskeletal disease (eg, osteoarthritis)	16 (33.3)	22 (44.0)	27 (55.1)	65 (44.2)
Neurological disease (eg, parkinsonism)	15 (31.3)	14 (28.0)	17 (34.7)	46 (31.3)
Mental disease (eg, anxiety)	16 (33.3)	15 (30.0)	12 (24.5)	43 (29.3)
Respiratory disease (eg, COPD)	7 (14.6)	5 (10.0)	10 (20.4)	22 (15.0)
Other disease (eg, kidney disease, cancer)	24 (50.0)	26 (52.0)	28 (57.1)	78 (53.1)
Level of education (number of participants), <i>n</i> (%)				
Less than secondary school	0	2 (4.0)	8 (16.3)	10 (6.8)
Secondary school graduation	6 (12.5)	8 (16.0)	5 (10.2)	19 (12.9)
Some postsecondary education	11 (22.9)	10 (20.0)	8 (16.3)	29 (19.7)
Postsecondary degree/diploma	31 (64.6)	30 (60.0)	28 (57.1)	89 (60.5)
Household income				
<\$20 000	0	1 (2.0)	4 (8.2)	5 (3.4)
\$20 000 or more, but less than \$50 000	5 (10.4)	10 (20.0)	12 (24.5)	27 (18.4)
\$50 000 or more, but less than \$100 000	14 (29.2)	21 (42.0)	21 (42.9)	56 (38.1)
\$100 000 or more, but less than \$150 000	13 (27.1)	10 (20.0)	5 (10.2)	28 (19.0)
\$150 000 or more	16 (33.3)	6 (12.0)	6 (12.2)	28 (19.0)
Unclear	0	2 (4.0)	1 (2.0)	3 (2.0)
Living alone, <i>n</i> (%)	5 (10.4)	6 (12.0)	18 (36.7)	29 (19.7)

Note: BMI = body mass index; COPD = chronic obstructive pulmonary disease; SD = standard deviations; Q1 = first quartile = median of the lower half of the data; Q3 = third quartile = median of the upper half of the data.

extreme values did not report any change in their physical health between visits.

### Variance Components Analysis

We summarized the results of the variance components analysis in [Supplementary Appendix Table S1](#). The proportions of variability between raters, different ages of participants, and between measurement times calculated from the variance components analysis were low for TUG and gait speed in participants aged 50–64 years and for single-leg stance in participants aged 75 years and older (ranging from 0% to 8.16%). The largest component of variance was residual variance (random error), the between-rater variance was very low and the between-participant variability was less than 50%, indicating that the TUG, gait speed, and single-leg stance may not be repeatable tests in those age groups.

### Absolute Reliability of Measures

Values for SEM and MDC for each of the PBMs are given in [Tables 2 and 3](#). For grip strength (mean and max) and chair rise, the SEMs

and MDCs were greater in participants aged 65–74 years than for participants in the younger and the older age groups. For TUG and gait speed, the SEMs were similar across the 3 age groups, and older participants had a lower MDC value. For single-leg stance (right leg, left leg, mean and max of 2 trials), the SEMs were larger in the older age groups; and participants in the 75 and older age group had the largest MDCs.

### Discussion

This study provides estimates of the relative and absolute reliabilities for grip strength, TUG, 4-m gait speed, chair-rise, and single-leg stance tests among community-dwelling older adults enrolled in the CLSA. Our findings indicate that for adults aged 50 and older, grip strength measures had excellent test–retest reliability; the TUG and single-leg stance tests had good reliability; and the gait speed and chair-rise tests had moderate reliability. In our age-stratified analyses, however, the TUG and gait speed, as well as the single-leg stance test, had poor test–retest reliabilities in the youngest and oldest age groups, respectively. Our study also provides some of the

**Table 2.** Relative and Absolute Reliabilities of Measures for All Participants

		Time I (baseline)	Time II (retest)	N	ICC (95% CI)	SEM	MDC <sub>90</sub>
Grip strength (mean, kg)	Mean (SD)	32.82 (11.71)	34.12 (12.38)	145	0.95 (0.92–0.97)	2.62	6.11
	Median (Q1, Q3)	30.47 (22.72, 42.63)	31.63 (22.96, 44.14)				
Grip strength (max, kg)	Mean (SD)	34.36 (12.12)	35.51 (12.71)	146	0.95 (0.93–0.97)	2.71	6.32
	Median (Q1, Q3)	31.50 (24.25, 43.93)	32.95 (24.18, 45.33)				
TUG (seconds)	Mean (SD)	10.47 (2.17)	10.07 (2.20)	147	0.80 (0.72–0.86)	0.97	2.26
	Median (Q1, Q3)	10.25 (9.16, 11.13)	9.50 (8.79, 10.97)				
4-m gait speed (m/s)	Mean (SD)	0.90 (0.16)	0.91 (0.15)	147	0.64 (0.54–0.73)	0.10	0.23
	Median (Q1, Q3)	0.91 (0.80, 1.01)	0.91 (0.83, 1.01)				
Chair rise (seconds)	Mean (SD)	12.23 (3.74)	10.95 (2.71)	144	0.64 (0.45–0.77)	2.24	5.23
	Median (Q1, Q3)	12.01 (9.45, 14.24)	10.67 (9.24, 12.49)				
Single-leg stance (right, seconds)	Mean (SD)	28.79 (23.65)	28.24 (25.06)	140	0.78 (0.70–0.84)	11.09	25.88
	Median (Q1, Q3)	18.96 (6.58, 60.00)	19.28 (4.08, 60.00)				
Single-leg stance (left, seconds)	Mean (SD)	29.33 (23.97)	27.74 (24.46)	140	0.78 (0.71–0.84)	11.24	26.23
	Median (Q1, Q3)	20.88 (5.78, 60.00)	16.91 (4.74, 60.00)				
Single-leg stance (mean, seconds)	Mean (SD)	29.06 (22.98)	27.99 (23.93)	140	0.82 (0.75–0.87)	9.75	22.75
	Median (Q1, Q3)	24.00 (7.16, 60.00)	19.44 (4.88, 60.00)				
Single-leg stance (max, seconds)	Mean (SD)	32.38 (23.26)	31.31 (24.38)	140	0.80 (0.72–0.85)	10.40	24.27
	Median (Q1, Q3)	28.28 (8.91, 60.00)	25.34 (5.71, 60.00)				

Note: ICC = intraclass correlation coefficient; SEM = standard error of measurement; MDC = minimal detectable change; TUG = Timed Up and Go; SD = standard deviation; Q1 = median of the lower half of the data; Q3 = median of the upper half of the data.

first estimates of MDC<sub>90</sub> values for each of the PBMs that can be used for interpreting change in physical function over time.

In our study, grip strength showed the highest test–retest reliability among the PBMs, with ICC values that were consistent with previous studies conducted in the community setting (9,29). Grip strength reflects overall muscle strength and is the simplest recommended method for assessing muscle function in clinical practice. However, methods for assessing grip strength have considerable variation, including whether the mean or maximum value, from 1, 2, or 3 trials, using the dominant hand only or either hand, is recorded (30). The results of our study support using 3 consecutive trials for the dominant hand, with a recording of either the mean or maximum value. The SEM of grip strength was about 2.5 kg, reflecting the estimated measurement error around a single assessment. Importantly, the MDC<sub>90</sub> of grip strength was about 6 kg, which is similar to estimates of meaningful changes in grip strength from anchor-based approaches (5.0–6.5 kg) (31).

The TUG test showed good test–retest reliability overall, similar to a previous study of community-dwelling older people ( $n = 20$ , mean age = 75 years) (32), but lower than another study in a community setting ( $n = 1200$ , mean age = 73 years; ICC = 0.93–0.99) (12), as well as a study that included patients with chronic conditions ( $n = 49$ , mean age = 50 years) (33). One recent study recruited healthy participants aged 50 years and older ( $n = 128$ , median age = 66 years) in Ireland and found the TUG test had moderate–good reliability (34). Our study reports a higher test–retest reliability than the previous study (34) (ICC: 0.80 vs 0.75) but yielded a higher MDC<sub>90</sub> value (2.26 vs 1.75 seconds). This can be explained by the larger standard deviations around the baseline scores in our study (2.17 vs 1.39 seconds), indicating that our data are spread out over a wider range. Of note, in our age-stratified analysis, we found that the TUG test had poor test–retest reliability (ICC = 0.38) for community-dwelling people aged 50–64 years. Unfortunately, it is difficult to compare these findings with previous studies as we are the first to examine test–retest reliability specifically for this age group. Our results suggest that the TUG may have limited utility in this “younger” adult age range in population-based studies.

The single-leg stance test had good test–retest reliability, which is consistent with previous studies in community-dwelling older people ( $n = 25$ , mean age = 72 years) (35) but lower than the other study in a community setting ( $n = 1200$ , mean age = 73 years; ICC = 0.93–0.99) (12) and studies including patients with chronic conditions ( $n = 71$ , mean age = 62 years) (36). Protocols for the single-leg stance test vary considerably in the literature with tests sometimes allowing the use of either leg or the preferred leg, from 1–5 trials, and scores recorded as the mean or maximum value of all trials, or the last of 2 trials (35,37). Our data indicate the mean value of the combined right leg trial and left leg trial had slightly better test–retest reliability (ICC = 0.82, MDC<sub>90</sub> = 22.75) than either trial on its own, or the maximum value of the 2 trials (ICC = 0.78–0.80).

In our study, the chair-rise and gait speed tests both had moderate test–retest reliabilities (ICC = 0.64 for both), values that are lower than those reported in previous studies of community-dwelling older people (ICC = 0.67–0.89) (11,13,14). It is worth noting that all 3 prior studies recruited participants aged 65 years and older; inclusion of younger participants in the current study likely decreased the between-participant variance and may explain our lower ICC values. If we consider our results only in participants aged 65 years and older, our ICC values improve slightly (0.67–0.70 vs 0.64). There are also several different protocol variations for both the chair-rise test and gait speed test which have different reported reliabilities (11,38). For example, previous studies reported the test–retest ICC for 3-m gait speed as 0.80 and 0.93 for the 5-m walk in older adults (6,7). The protocols with the best reliability for both chair-rise and gait speed tests need to be further investigated.

The ICC reflects the proportion of between-participant variance on the total variance (23). In our study, the major variances for the PBMs with poor reliability in some of the age-stratified analyses are from error (57%–72%). Although we expected that raters would play an important role in the source of the variation (34), we did not find such effects in our variance component analysis. Our variance component analysis results indicate that different raters collecting data over time may not be the major source of variance for the measurement results if clear standard operating procedures are followed,

**Table 3.** Relative and Absolute Reliabilities of Measures for Different Age Groups

	Time I (baseline)	Time II (retest)	N	ICC (95% CI)	SEM	MDC <sub>90</sub>
<i>Grip strength (mean, kg)</i>						
50–64 years						
Mean (SD)	34.78 (12.32)	36.07 (12.38)	48	0.97 (0.94–0.98)	2.13	4.97
Median (Q1, Q3)	30.57 (26.34, 44.36)	31.87 (27.38, 44.92)				
65–74 years						
Mean (SD)	35.33 (12.28)	37.48 (13.04)	49	0.93 (0.84–0.96)	3.25	7.58
Median (Q1, Q3)	33.7 (23.87, 46.9)	36.63 (25.12, 48.58)				
75+ years						
Mean (SD)	28.29 (9.08)	34.78 (12.32)	48	0.96 (0.92–0.98)	1.82	4.25
Median (Q1, Q3)	27.63 (19.88, 35.75)	28.33 (19.88, 38.1)				
<i>Grip strength (max, kg)</i>						
50–64 years						
Mean (SD)	36.65 (12.63)	37.56 (12.84)	48	0.97 (0.95–0.98)	2.19	5.11
Median (Q1, Q3)	31.65 (27.68, 46.18)	32.7 (28.8, 46)				
65–74 years						
Mean (SD)	37.08 (12.79)	39.08 (13.23)	49	0.91 (0.84–0.95)	3.84	8.96
Median (Q1, Q3)	35 (25.05, 48.4)	38 (26.9, 50.55)				
75+ years						
Mean (SD)	29.41 (9.28)	36.65 (12.63)	49	0.96 (0.93–0.98)	1.86	4.34
Median (Q1, Q3)	29.5 (21.05, 37)	29.3 (20.65, 39.55)				
<i>TUG (seconds)</i>						
50–64 years						
Mean (SD)	9.46 (1.52)	8.93 (1.01)	48	0.38 (0.12–0.59)	1.2	2.8
Median (Q1, Q3)	9.24 (8.68, 10.1)	8.89 (8.34, 9.48)				
65–74 years						
Mean (SD)	10.4 (2.47)	10.02 (2.65)	50	0.85 (0.75–0.91)	0.96	2.24
Median (Q1, Q3)	10.27 (9.24, 10.81)	9.49 (8.71, 10.62)				
75+ years						
Mean (SD)	11.53 (1.92)	11.24 (1.98)	49	0.79 (0.66–0.88)	0.88	2.05
Median (Q1, Q3)	11.31 (10.22, 12.52)	10.97 (9.69, 13)				
<i>4-m gait speed (m/s)</i>						
50–64 years						
Mean (SD)	0.95 (0.13)	0.98 (0.11)	48	0.33 (0.05–0.55)	0.11	0.26
Median (Q1, Q3)	0.97 (0.91, 1.05)	0.96 (0.89, 1.06)				
65–74 years						
Mean (SD)	0.91 (0.17)	0.93 (0.15)	50	0.67 (0.48–0.80)	0.1	0.23
Median (Q1, Q3)	0.89 (0.82, 1.02)	0.92 (0.86, 1.02)				
75+ years						
Mean (SD)	0.84 (0.15)	0.83 (0.17)	49	0.69 (0.50–0.81)	0.08	0.19
Median (Q1, Q3)	0.84 (0.74, 0.92)	0.83 (0.72, 0.92)				
<i>Chair rise (seconds)</i>						
50–64 years						
Mean (SD)	11.75 (4.74)	9.94 (2.33)	48	0.55 (0.26–0.74)	3.18	7.42
Median (Q1, Q3)	10.49 (8.86, 13.53)	9.64 (8.51, 11.52)				
65–74 years						
Mean (SD)	11.59 (3.11)	10.36 (2.42)	48	0.70 (0.38–0.85)	1.7	3.97
Median (Q1, Q3)	11.75 (9.45, 13.3)	10.30 (9.36, 12.24)				
75+ years						
Mean (SD)	13.33 (2.94)	12.56 (2.66)	48	0.67 (0.47–0.80)	1.69	3.94
Median (Q1, Q3)	13.68 (10.64, 15.32)	12.44 (10.28, 13.75)				
<i>Single-leg stance (right, seconds)</i>						
50–64 years						
Mean (SD)	46.52 (19.57)	49.12 (19.26)	48	0.78 (0.64–0.87)	9.18	21.42
Median (Q1, Q3)	60.00 (27.08, 60)	60.00 (41.55, 60)				
65–74 years						
Mean (SD)	25.16 (21.72)	24.68 (23.27)	46	0.73 (0.55–0.84)	11.29	26.34
Median (Q1, Q3)	17.41 (5.5, 47.71)	12.38 (4.44, 58.69)				
75+ years						
Mean (SD)	14 (16.87)	10.4 (14.48)	46	0.30 (0.02–0.54)	14.04	32.76
Median (Q1, Q3)	8.89 (2.33, 15.6)	4.06 (2.09, 14.13)				
<i>Single-leg stance (left, seconds)</i>						
50–64 years						
Mean (SD)	46.01 (19.61)	44.82 (21.43)	48	0.73 (0.56–0.84)	10.19	23.78
Median (Q1, Q3)	60.00 (27.2, 60)	60.00 (21.79, 60)				

Table 3. Continued

	Time I (baseline)	Time II (retest)	N	ICC (95% CI)	SEM	MDC <sub>90</sub>
65–74 years						
Mean (SD)	29.36 (22.97)	27.15 (23.67)	46	0.76 (0.61–0.86)	11.25	26.25
Median (Q1, Q3)	21.07 (6.96, 60)	18.80 (5.29, 60.00)				
75+ years						
Mean (SD)	11.89 (15.45)	10.86 (14.4)	46	0.38 (0.10–0.60)	12.17	28.4
Median (Q1, Q3)	5.92 (3.15, 13.14)	4.84 (1.81, 13.62)				
<i>Single-leg stance (mean, seconds)</i>						
50–64 years						
Mean (SD)	46.26 (18.93)	46.97 (19.31)	48	0.78 (0.64–0.87)	8.88	20.72
Median (Q1, Q3)	60.00 (27.35, 60)	60.00 (33.13, 60)				
65–74 years						
Mean (SD)	27.26 (21.21)	25.91 (22.07)	46	0.80 (0.67–0.89)	9.49	22.14
Median (Q1, Q3)	25.97 (7.69, 40.2)	16.44 (6.23, 44.57)				
75+ years						
Mean (SD)	12.95 (15.01)	10.63 (14.05)	46	0.39 (0.11–0.61)	11.72	27.35
Median (Q1, Q3)	7.25 (3.54, 15.31)	3.91 (2.31, 13.08)				
<i>Single-leg stance (max, seconds)</i>						
50–64 years						
Mean (SD)	48.63 (17.08)	50.17 (18.04)	48	0.76 (0.60–0.86)	8.37	19.53
Median (Q1, Q3)	60.00 (39.87, 60)	60.00 (42.78, 60)				
65–74 years						
Mean (SD)	31.62 (22.39)	30.67 (23.27)	46	0.83 (0.71–0.90)	9.23	21.54
Median (Q1, Q3)	28.41 (9.87, 60)	24.35 (8.31, 60)				
75+ years						
Mean (SD)	16.21 (17.77)	12.69 (14.86)	46	0.31 (0.03–0.55)	14.76	34.44
Median (Q1, Q3)	9.39 (5, 18.16)	5.59 (3.09, 17.94)				

Note: ICC = intraclass correlation coefficient; SEM = standard error of measurement; MDC = minimal detectable change; TUG = Timed Up and Go; SD = standard deviation; Q1 = median of the lower half of the data; Q3 = median of the upper half of the data.

such as in the CLSA. The poor reliabilities observed in some of the age-stratified analyses may be due to a more homogeneous healthy population compared with older age groups or populations from clinical settings for the chair-rise test and gait speed. For the single-leg stance test, the exclusion of participants with high fall risk from performing the test may have resulted in a more homogeneous population for the 75 and older age group.

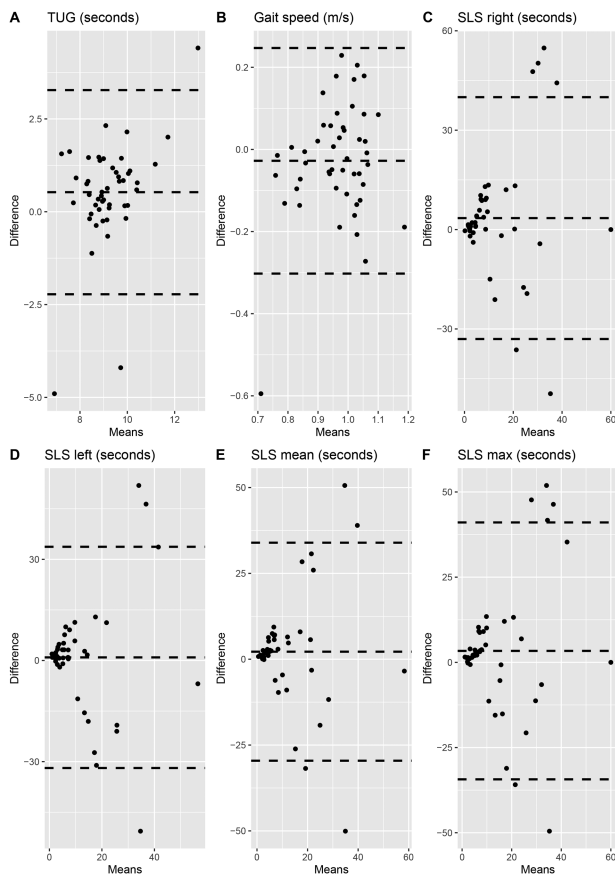
The age-stratified MDC<sub>90</sub> values for the PBMs were consistent with our relative reliability results. The MDC<sub>90</sub> for TUG and gait speed, and the single-leg stance test, were largest in the youngest (50–64 years) and oldest (75+ years) age groups, respectively, because of the lower ICC values in these age groups. As such the MDC<sub>90</sub> values for those tests are larger than previous estimates of clinically meaningful change for some of the PBMs (39,40). Of note, the MDC<sub>90</sub> estimates for the single-leg stance test in the oldest age group were larger than the mean and median values of the test scores (Table 3). These results are consistent with a previous study conducted in community-dwelling older adults ( $n = 25$ , mean age = 72 years) (35) and suggest that the utility of the single-leg stance test in participants older than 75 years needs further investigation. Overall, these findings highlight the importance of considering age when interpreting change in PBMs over time among community-dwelling older adults aged 50 and older. Importantly, the MDC<sub>90</sub> values reported here may not apply to specific clinical populations and those from groups with more narrow age ranges.

This is the first study to provide test–retest reliability data for the CLSA PBMs in middle-aged and older adults. In addition, we report age-stratified results and explore sources of potential variability for

the PBMs with poor reliability using variance component analysis. Nevertheless, this study had several limitations. First, the number of participants in some age groups is slightly smaller (eg, 46 for single-leg stance tests in the older 2 groups) than the recommended sample size for reliability studies ( $n = 50$ ), which might affect the precision of the point estimates. Second, we approached consecutive participants from a single CLSA data collection site using a convenience sampling method and therefore may have unintentionally introduced selection bias; those with more health/physical limitations may have been less willing to participate in the study compared to higher functioning community-dwelling people. Third, we did not account for cognitive status, which may be another source of variance (41). Fourth, this study was a substudy within the CLSA whereby these measures were administered manually; use of automated assessments or video recordings may have improved our findings (42,43), however would have less external validity for these methods or devices are not commonly used.

## Conclusions

Among community-dwelling Canadians aged 50 and older enrolled in the CLSA, overall test–retest reliability values for measures of grip strength, TUG, single-leg stance, gait speed, and chair rise were moderate to excellent. The TUG test and gait speed had poor reliability in people aged 50–64 years, and the single-leg stance had poor reliability in people aged 75 years and older. MDC values presented in this article can be used to help interpret changes in physical function over time.



**Figure 1.** Bland–Altman plot for measures that exhibited poor reliability: **A** and **B** among participants aged 50–64 years and **C–F** among participants aged 75 years and older. TUG = Timed Up and Go; SLS = single-leg stance.

**Supplementary Material**

Supplementary data are available at *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences* online.

**Funding**

M.K.B. holds a Tier 2 Canada Research Chair in Mobility, Aging, and Chronic Disease. This research was conducted in collaboration with the Canadian Longitudinal Study on Aging (CLSA), specifically the Hamilton Site Data Collection Centre. Funding for the CLSA is provided by the Government of Canada through the Canadian Institutes of Health Research (CIHR) under grant reference: LSA 94473 and the Canada Foundation for Innovation. The CLSA is led by P.R., C.W., and S.K. P.R. holds the Raymond and Margaret Labarge Chair in Research and Knowledge, is the Director of the McMaster Institute for Research on Aging and the Labarge Centre for Mobility in Aging, and holds a Tier 1 Canada Research Chair in Geroscience.

**Acknowledgments**

M.K.B., A.K., L.G., and P.R. conceived the study and secured funding for this study. Q.H. conducted the data analysis and drafted the initial manuscript supervised by M.K.B. C.D., G.S., C.B., and A.M. were involved in data collection and study design. M.K.B., P.R., L.G., and J.M. helped with data analysis and results interpretation. C.W. and S.K. gave critical feedback on the manuscript. All authors approved the final manuscript.

**Conflict of Interest**

None declared.

**References**

1. Karunanathan S, Moodie EEM, Bergman H, et al. The association between physical function and proximity to death in older adults: a multilevel analysis of 4,150 decedents from the Cardiovascular Health Study. *Ann Epidemiol.* 2019;35:59–65.e5. doi:10.1016/j.annepidem.2019.04.005
2. Trombetti A, Reid KF, Hars M, et al. Age-associated declines in muscle mass, strength, power, and physical performance: impact on fear of falling and quality of life. *Osteoporos Int.* 2016;27:463–471. doi:10.1007/s00198-015-3236-5
3. Hardy SE, Kang Y, Studenski SA, Degenholtz HB. Ability to walk 1/4 mile predicts subsequent disability, mortality, and health care costs. *J Gen Intern Med.* 2011;26:130–135. doi:10.1007/s11606-010-1543-2
4. O’Neill D, Forman DE. The importance of physical function as a clinical outcome: assessment and enhancement. *Clin Cardiol.* 2020;43:108–117. doi:10.1002/clc.23311
5. Seeman TE, Charpentier PA, Berkman LF, et al. Predicting changes in physical performance in a high-functioning elderly cohort: MacArthur studies of successful aging. *J Gerontol.* 1994;49:M97–108. doi:10.1093/geronj/49.3.m97
6. Nevitt MC, Cummings SR, Kidd S, Black D. Risk factors for current nonsyncopal falls: a prospective study. *J Am Med Assoc.* 1989;261:2663–2668. doi:10.1001/jama.1989.03420180087036
7. Jette AM, Jette DU, Ng J, Plotkin DJ, Bach MA. Are performance-based measures sufficiently reliable for use in multicenter trials? Musculoskeletal Impairment (MSI) Study Group. *J Gerontol A Biol Sci Med Sci.* 1999;54:M3–M6. doi:10.1093/gerona/54.1.m3
8. 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. doi:10.1111/j.1532-5415.1991.tb01616.x
9. Bohannon RW, Schaubert KL. Test–retest reliability of grip-strength measures obtained over a 12-week interval from community-dwelling elders. *J Hand Ther.* 2005;18:426–428. doi:10.1197/j.jht.2005.07.003
10. 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. doi:10.1093/gerona/55.2.m70
11. Kim HJ, Park I, Lee HJ, Lee O. The reliability and validity of gait speed with different walking pace and distances against general health, physical function, and chronic disease in aged adults. *J Exerc Nutrition Biochem.* 2016;20:46–50. doi:10.20463/jenb.2016.09.20.3.7
12. Lin MR, Hwang HF, Hu MH, Wu HD, Wang YW, Huang FC. Psychometric comparisons of the timed up and go, one-leg stand, functional reach, and Tinetti balance measures in community-dwelling older people. *J Am Geriatr Soc.* 2004;52:1343–1348. doi:10.1111/j.1532-5415.2004.52366.x
13. Tiedemann A, Shimada H, Sherrington C, Murray S, Lord S. The comparative ability of eight functional mobility tests for predicting falls in community-dwelling older people. *Age Ageing.* 2008;37:430–435. doi:10.1093/ageing/afn100
14. Jette AM, Jette DU, Ng J, et al. Are performance-based measures sufficiently reliable for use in multicenter trials? Musculoskeletal Impairment (MSI) Study Group. *J Gerontol A Biol Sci Med Sci.* 1999; 54(1):M3–M6. doi:10.1093/gerona/54.1.m3
15. Goldberg A, Schepens S. Measurement error and minimum detectable change in 4-meter gait speed in older adults. *Aging Clin Exp Res.* 2011;23:406–412. doi:10.1007/BF03325236
16. Goldberg A, Chavis M, Watkins J, Wilson T. The five-times-sit-to-stand test: validity, reliability and detectable change in older females. *Aging Clin Exp Res.* 2012;24:339–344. doi:10.1007/BF03325265
17. Ferrucci L, Cooper R, Shardell M, Simonsick EM, Schrack JA, Kuh D. Age-related change in mobility: perspectives from life course epidemiology



- and geroscience. *J Gerontol A Biol Sci Med Sci*. 2016;71:1184–1194. doi:10.1093/gerona/glw043
18. Raina PS, Wolfson C, Kirkland SA, et al. The Canadian longitudinal study on aging (CLSA). *Can J Aging*. 2009;28:221–229. doi:10.1017/S0714980809990055
  19. Raina P, Wolfson C, Kirkland S, et al. Cohort profile: the Canadian Longitudinal Study on Aging (CLSA). *Int J Epidemiol*. 2019;48:1752–1753j. doi:10.1093/ije/dyz173
  20. Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluations. *J Hand Surg Am*. 1984;9:222–226. doi:10.1016/s0363-5023(84)80146-x
  21. Wellmon R. Chapter 32—gait assessment and training. In: Cameron MH, Monroe LG, eds. *Physical Rehabilitation Examination, Evaluation and Intervention: An Evidence-Based Approach*. Philadelphia: W.B. Saunders; 2007:844–876. doi:10.1016/B978-0-7216-0361-2.X5001-7
  22. Bohannon RW, Bubela DJ, Magasi SR, Wang YC, Gershon RC. Sit-to-stand test: performance and determinants across the age-span. *Isokinet Exerc Sci*. 2010;18:235–240. doi:10.3233/IES-2010-0389
  23. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–163. doi:10.1016/j.jcm.2016.02.012
  24. Dontje ML, Dall PM, Skelton DA, Gill JMR, Chastin SFM; Seniors USP Team. Reliability, minimal detectable change and responsiveness to change: indicators to select the best method to measure sedentary behaviour in older adults in different study designs. *PLoS One*. 2018;13:e0195424. doi:10.1371/journal.pone.0195424
  25. Andre Schuetzenmeister FD. *VCA: Variance Component Analysis. R package version 1.4.3*. 2020. <https://cran.r-project.org/web/packages/VCA/VCA.pdf>
  26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310. doi:10.1016/S0140-6736(86)90837-8
  27. Chen S, Wolf SL, Zhang Q, Thompson PA, Winstein CJ. Minimal detectable change of the actual amount of use test and the motor activity log: the EXCITE Trial. *Neurorehabil Neural Repair*. 2012;26:507–514. doi:10.1177/1545968311425048
  28. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med*. 2002;21:1331–1335. doi:10.1002/sim.1108
  29. Bohannon RW. Test–retest reliability of measurements of hand-grip strength obtained by dynamometry from older adults: a systematic review of research in the PubMed database. *J Frailty Aging*. 2017;6:83–87. doi:10.14283/jfa.2017.8.
  30. Roberts HC, Denison HJ, Martin HJ, et al. A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach. *Age Ageing*. 2011;40:423–429. doi:10.1093/ageing/afr051
  31. Bohannon RW. Minimal clinically important difference for grip strength: a systematic review. *J Phys Ther Sci*. 2019;31:75–78. doi:10.1589/jpts.31.75
  32. Bohannon RW, Schaubert K. Long-term reliability of the timed up-and-go test among community-dwelling elders. *J Phys Ther Sci*. 2005;17:93–96. doi:10.1589/jpts.17.93
  33. Collado-Mateo D, Domínguez-Muñoz FJ, Adsuar JC, Merellano-Navarro E, Olivares PR, Gusi N. Reliability of the timed up and go test in fibromyalgia. *Rehabil Nurs*. 2018;43:35–39. doi:10.1002/rnj.307
  34. Donoghue OA, Savva GM, Börsch-Supan A, Kenny RA. Reliability, measurement error and minimum detectable change in mobility measures: a cohort study of community-dwelling adults aged 50 years and over in Ireland. *BMJ Open*. 2019;9:e030475. doi:10.1136/bmjopen-2019-030475
  35. Goldberg A, Casby A, Wasielewski M. Minimum detectable change for single-leg-stance-time in older adults. *Gait Posture*. 2011;33:737–739. doi:10.1016/j.gaitpost.2011.02.020
  36. Ortega-Perez de Villar L, Martínez-Olmos FJ, Junque-Jimenez A, et al. Test–retest reliability and minimal detectable change scores for the short physical performance battery, one-legged standing test and timed up and go test in patients undergoing hemodialysis. *PLoS One*. 2018;13:e0201035. doi:10.1371/journal.pone.0201035
  37. Bohannon RW. Single limb stance times: a descriptive meta-analysis of data from individuals at least 60 years of age. *Top Geriatr Rehabil*. 2006;22:70–77. doi:10.1097/00013614-200601000-00010
  38. Mehmet H, Yang AWH, Robinson SR. What is the optimal chair stand test protocol for older adults? A systematic review. *Disabil Rehabil*. 2020;42:2828–2835. doi:10.1080/09638288.2019.1575922
  39. Kon SS, Canavan JL, Nolan CM, et al. The 4-metre gait speed in COPD: responsiveness and minimal clinically important difference. *Eur Respir J*. 2014;43:1298–1305. doi:10.1183/09031936.00088113.
  40. Wright AA, Cook CE, Baxter GD, Dockerty JD, Abbott JH. A comparison of 3 methodological approaches to defining major clinically important improvement of 4 performance measures in patients with hip osteoarthritis. *J Orthop Sports Phys Ther*. 2011;41:319–327. doi:10.2519/jospt.2011.3515
  41. Ibrahim A, Singh DKA, Shahar S. ‘Timed Up and Go’ test: age, gender and cognitive impairment stratified normative values of older adults. *PLoS One*. 2017;12:e0185641. doi:10.1371/journal.pone.0185641
  42. Collado-Mateo D, Madeira P, Dominguez-Munoz FJ, Villafaina S, Tomas-Carus P, Parraca JA. The automatic assessment of strength and mobility in older adults: a test–retest reliability study. *Medicina (Kaunas)*. 2019;55:270. doi:10.3390/medicina55060270.
  43. Sprint G, Cook DJ, Weeks DL. Toward automating clinical assessments: a survey of the timed up and go. *IEEE Rev Biomed Eng*. 2015;8:64–77. doi:10.1109/RBME.2015.2390646