BMC Geriatrics

RESEARCH Open Access

Reliability of wearable sensors single and dual-task timed up and go test among community-dwelling older women

Xiao Yu[1*] and Roxana Dev Omar Dev[1*]

Abstract

Background Approximately one third of community-dwelling older adults fall each year, and older women experience roughly twice the rate and number of falls as men. The timed up and go test (TUG) is a widely used to assess mobility and fall risk. However, its ability to detect subtle gait impairments may be constrained in high-functioning older adults.

Objective This study aimed to evaluate the reliability of gait parameters measured by wearable inertial sensors during single-task timed up and go test (ST-TUG) and dual-task timed up and go test (DT-TUG) in community-dwelling older women, to assess their usefulness for fall-risk screening.

Methods Twenty-five cognitively healthy women (mean age 70.09 ± 3.52 years) completed ST-TUG and DT-TUG within a two-day interval. Wearable inertial sensors yielded 21 parameters, including completion time, spatiotemporal measures, coefficients of variation, and turn-related metrics. Reliability was assessed using intraclass correlation coefficients (ICC), the standard error of measurement (SEM and SEM%), and the minimal detectable change at 95% confidence (MDC_{95}).

Results Gait spatiotemporal parameters demonstrated good to excellent test–retest reliability (ICC ≥ 0.75) under both ST and DT conditions, whereas gait variability and turn-related measures showed poor reliability (ICC < 0.50). Measurement error was slightly higher under dual-task conditions. The DT-TUG appeared more sensitive to gait disturbances relevant to fall risk.

Conclusions Combining wearable sensors with ST-TUG and DT-TUG provides reliable, repeatable measurement of core timing/pace metrics in older women. Although variability and turn-related measures exhibit limited short-interval reliability, DT-TUG may enhance early fall-risk screening when paired with detailed gait analysis.

Keywords Wearable sensors, Timed up and go test, Fall risk, Older adults, Reliability

*Correspondence:
Xiao Yu
gs61542@student.upm.edu.my
Roxana Dev Omar Dev
rdod@upm.edu.my
[1]Department of Sports Studies, Universiti Putra Malaysia UPM, Serdang, Selangor, Malaysia

Introduction

With global population ageing, falls among older adults have become a major public health concern. The World Health Organization (WHO) reports that older adults are particularly susceptible to serious injury and death due to falls, with risk increasing with age and being especially high among older women [32]. Timely assessment of fall risk is therefore essential for implementing effective prevention strategies.

The timed up and go test (TUG) test is a modification of the "get up and go" test in which a participant stands up from a chair, walks forward three metres, turns around, and walks back to the chair and sits back down [20]. TUG test is widely used to evaluate functional mobility and fall risk in older adults [34]. Among adults aged ≥ 65 years, faster TUG completion time reflects superior functional mobility, and a threshold of ≥ 13.5 s has been proposed to indicate higher fall risk [21]. However, a review showed that the TUG test was not a significant predictor of falls [2]. This is consistent with the results of another review, which demonstrated that its capacity to predict falls is extremely limited in older adults and is not adequate as a standalone predictive tool for identifying fall risk [3]. Moreover, a ceiling effect has been observed limiting sensitivity to subtle differences in fall risk or early functional decline [23]. These limitations underscore the need for more specific and sensitive gait parameters to accurately characterise walking ability in this population.

Gait parameters have been shown to play a significant role in clinical rehabilitation, serving as quantitative data for the assessment of mobility function, the diagnosis of diseases, and the prediction of fall risk. The advent of wearable sensor technology has precipitated a paradigm shift in the field, with its application in the evaluation of the ambulatory capacity of older individuals becoming a salient feature. It is capable of recording a number of fundamental walking parameters, in addition to capturing more nuanced details of gait and turning processes, which are important in predicting fall risk. Greene et al. assessed 349 community-dwelling older adults using wearable sensors and demonstrated that the wearable sensor TUG test (WTUG) outperformed manual timing for discriminating fall risk [8]. Similarly, Choi et al. developed a wearable sensor–based predictive model to support clinicians in estimating fall risk in older adults [5].

However, it is inevitable that older adults will encounter situations in their lives where they perform dual tasks, for instance, talking while walking or searching for items in their pocket during household activities. Even simple walking is accompanied by perceptual and cognitive demands imposed by complex environments [1]. Performance during dual-tasking is contingent on the ability to allocate attentional resources between the two tasks when they interfere with each other. Dual-task paradigms include a cognitive–motor paradigm (e.g., mental arithmetic, reverse sequencing, verbal fluency) and a motor–motor paradigm (e.g., performing the test while holding a glass of water or carrying an object). These paradigms place different demands on the nervous system: cognitive–motor tasks primarily load working memory, sustained attention, and executive control, whereas motor–motor tasks add concurrent motor control requirements. Siu et al. has noted that flexible allocation of attention between the two tasks when walking in dual-tasking conditions may be a dual-task performance [25]. A substantial body of research has showed that older adults at higher fall risk exhibit greater gait variability under dual-task conditions [15, 19]. This finding suggests that the assessment of dual-tasking gait parameters may have significant value in predicting and screening for fall risk. A prospective study reported that following 12 months of observation, DT-TUG more accurately predicted future falls than the 30-second sit-to-stand test, the five-times sit-to-stand test, and ST-TU [29].

To date, there has been a paucity of studies that have combined wearable sensors to systematically assess the reliability of ST-TUG and DT-TUG. Furthermore, there is a lack of multidimensional quantitative metrics to evaluate the results of both tests. In the present study, we used wearable inertial sensors to conduct ST-TUG and DT-TUG separately. Because the magnitude and locus of interference are task-dependent, we purposefully adopted a cognitive–motor DT-TUG using serial subtraction by 3 s to impose a controlled working-memory load that is closely linked to executive aspects of gait regulation (particularly during turning and step-to-step control). Therefore, the reliability profile reported herein should be interpreted within this cognitive load context. The aim of this study was to determine the reliability of specific gait parameters in the ST-TUG and DT-TUG using specific inertial-measurement-unit in community-dwelling older women, quantified by intraclass correlation coefficients (ICC), standard error of measurement (SEM and SEM%), and the minimal detectable change at 95% confidence (MDC_{95}), with a view to assessing their potential value for fall-risk screening.

Methods

Study design

In this study, TUG performance was recorded under two conditions: ST-TUG and DT-TUG (cognitive-motor). An inertial measurement unit (IMU; MPU6050) was used to capture motion signals during both tests. Assessments were conducted in a standard home environment. Participants completed ST-TUG and DT-TUG on two consecutive days, separated by 24–48 h. Test location, surroundings, and procedures were kept identical across sessions to maximise comparability. From the sensor

recordings, 21 gait parameters were retained for reliability analysis, covering temporal, spatial, angular-velocity, and turn-related domains [26]. After each session, raw data were uploaded to the Kinesis QTUG™ system for extraction of spatiotemporal metrics and computation of fall-risk outputs.

Participants

Twenty-five community-dwelling older women from Wuhan, China, were enrolled. Eligibility was determined via screening before testing. Inclusion criteria were: age ≥ 65 years; no significant cognitive impairment (Montreal Cognitive Assessment, MoCA ≥ 26); no history of falls in the preceding 6 months; no severe psychiatric or musculoskeletal disorders; and no recent history of medical interventions or rehabilitation. All participants received study information and provided written informed consent.

The experiment was conducted in accordance with the approval and clearance of the Ethics Committee of Universiti Putra Malaysia (Proceedings JKEUPM-2023-373). Before testing, participants completed the Berg Balance Scale (BBS) and Activities-Specific Balance Confidence (ABC) Scale. In this study, the Chinese version of the Berg Balance Scale, which was designed by Jin Dongmei, was utilised as a tool to evaluate balance function in older adults, this version of the scale has been demonstrated to possess high validity, reliability and sensitivity when employed for the purpose of assessing balance function. In this study, the Chinese version of the ABC scale, which was designed by Qiang Guan, was utilised as a tool to evaluate the balance confidence of older adults [10]. This scale has been demonstrated to possess excellent reliability and validity, making it a suitable instrument for the assessment of balance confidence in older adults in mainland China [9]. Mean scores were BBS = 43.78 and ABC = 65.26, indicating a moderate fall risk and low balance confidence in this cohort. Participant characteristics are summarised in Table 1.

Table 1 Characteristics of subjects ($N = 25$) mean (Standard Deviation)

Characteristics	Mean (Standard Deviation)
Age (years)	70.09 (3.52)
Body mass (kg)	66.87 (3.66)
Body height (cm)	156.61 (4.47)
BMI (kg/m^2)	27.27 (1.05)
MoCA	26.87 (0.81)
BBS	43.78 (1.44)
ABC Scale	65.26 (2.22)

BMI Body Mass Index, *MoCA* Montreal Cognitive Assessment, *BBS* Berg Balance Scale, *ABC Scale* Activities-Specific Balance Confidence Scale

Procedure

This study utilised a 6-axis IMU comprising a tri-axial accelerometer and gyroscope to capture gait signals at a sampling frequency of 100 Hz. For each session, a single sensor was securely fastened to the anteromedial aspect of the shank using an elastic band. Participants were asked to refrain from strenuous exercise or therapeutic interventions during the study period. A trained surveyor assisted the subjects in wearing the sensors and completing ST-TUG and DT-TUG, maintaining the same order for both tests.

Two TUG conditions were recorded:

Single-task timed up and go test (ST-TUG): Participants wore the sensor, rose from a chair, walked 3 m to a marker, turned around it, returned to the chair, and sat down.

Dual-task timed up and go test (DT-TUG, cognitive–motor): While performing the TUG sequence, participants concurrently answered simple mental-arithmetic prompts delivered by the assessor (e.g., serial subtraction by 3 from 100) and were instructed to maintain a comfortable, steady walking pace.

Each condition was completed twice (test and retest) on two consecutive days with a 24–48 h interval. Testing location, environmental conditions, and procedures were held constant across sessions. After each trial, spatiotemporal gait variables were extracted and synchronised to the Kinesis QTUG™ system.

Statistical analysis

The present study evaluated the reliability of the measurements according to the established procedure proposed by Weir [30]. The data were analysed using SPSS 26.0. Data completeness was checked prior to analysis. Following the determination that there were no outliers, the subsequent step in the analysis was performed. Descriptive statistics summarised participant characteristics and outcome measures. Distributional assumptions were examined using the Shapiro–Wilk test; $p \geq 0.05$ was interpreted as no significant deviation from normality [7]. When the normality assumption was not met, within-subject test-retest comparisons were performed using the paired non-parametric Wilcoxon signed-rank test. All tests were two-sided with $\alpha = 0.05$. The mean and standard deviation of the results of ST-TUG and DT-TUG were calculated (a total of 21 gait parameters including time, space, angular velocity and turning parameters). Test–retest reliability was quantified using intraclass correlation coefficients (ICC) based on a two-way mixed-effects, absolute-agreement, single-measure model with 95% confidence intervals (CI). In the context of reliability assessment, an ICC ≥ 0.90 is indicative of very high reliability, 0.75 ≤ ICC < 0.9 indicates good reliability, 0.5 ≤ ICC < 0.75 indicates moderate reliability,

and ICC < 0.5 indicates poor reliability [14]. In order to quantify the random measurement error of each outcome metric based on the sample standard deviation (SD) and the corresponding ICC, the formula was used to calculate standard error of measurement (SEM). The SEM is the standard deviation of errors of measurement that are associated with test scores from a particular group of examinees [11]. The SEM was calculated as $\mathrm{SEM} = \mathrm{SD} \times \sqrt{1 - \mathrm{ICC}}$ [33]. Furthermore, the minimum detectable change at 95% confidence level (MDC_{95}) is calculated, and the MDC_{95} value represents the amount of change required for a particular measurement metric to exceed the measurement error based on a pre-determined confidence threshold [28]. MDC_{95} was calculated as $\mathrm{MDC}_{95} = 1.96 \times \sqrt{2} \times \mathrm{SEM}$ [27].

Table 2 Comparison of two measurements of gait parameters ($N = 25$)

Gait Parameters	Task Type	First measurement	Second measurement
Time to Complete Test (s)	ST	8.69 (0.77)	8.89 (0.97)
	DT	11.68 (0.92)	11.70 (0.93)
Walk Time (s)	ST	5.31 (0.58)	5.41 (0.62)
	DT	7.92 (0.70)	7.93 (0.72)
Mean Stance Time (s)	ST	0.64 (0.02)	0.64 (0.03)
	DT	0.75 (0.02)	0.75 (0.02)
Mean Step Time (s)	ST	0.60 (0.02)	0.61 (0.04)
	DT	0.71 (0.04)	0.72 (0.07)
Mean Stride Time (s)	ST	1.20 (0.05)	1.20 (0.06)
	DT	1.45 (0.08)	1.44 (0.08)
Mean Stride Length (cm)	ST	67.49 (1.71)	68.28 (1.46)
	DT	73.51 (1.55)	74.11 (1.80)
Mean Stride Velocity (cm/s)	ST	56.67 (1.41)	56.53 (1.79)
	DT	50.67 (1.09)	50.60 (1.05)
Number of Gait Cycles	ST	3.32 (0.48)	3.44 (0.65)
	DT	4.52 (0.77)	4.48 (0.51)
Number of Steps	ST	10.28 (0.84)	10.32 (0.80)
	DT	13.80 (1.08)	13.96 (0.84)
Cadence (steps/min)	ST	100.44 (4.10)	99.88 (5.69)
	DT	81.88 (2.57)	81.80 (2.43)
Double Support % of Gait Cycle (%)	ST	38.43 (1.28)	38.34 (1.45)
	DT	44.44 (1.89)	44.01 (2.98)
Coefficient of Variability Stride Time (%)	ST	2.87 (0.32)	2.74 (0.33)
	DT	4.23 (0.40)	4.34 (0.64)
Coefficient of Variability Step Time (%)	ST	3.37 (0.27)	3.17 (0.37)
	DT	4.66 (0.45)	4.67 (0.46)
Coefficient of Variability Stride Velocity (%)	ST	4.11 (0.57)	3.93 (0.37)
	DT	5.25 (0.51)	5.40 (0.45)
Coefficient of Variability Stride Length (%)	ST	2.73 (0.21)	2.94 (0.48)
	DT	3.79 (0.47)	4.04 (0.55)
Coefficient of Variability Single Support (%)	ST	3.37 (0.20)	3.40 (0.48)
	DT	4.93 (0.50)	4.86 (0.44)
Turn Mid-Point Time (s)	ST	1.94 (0.12)	1.93 (0.14)
	DT	2.69 (0.13)	2.69 (0.20)
Turn Time (s)	ST	2.13 (0.17)	2.24 (0.23)
	DT	2.92 (0.31)	2.87 (0.25)
Return from Turn Time (s)	ST	1.76 (0.08)	1.78 (0.11)
	DT	2.51 (0.11)	2.52 (0.16)
Turning Magnitude (degrees/sec)	ST	96.00 (4.07)	97.48 (6.96)
	DT	87.72 (2.82)	87.64 (2.66)
Number Strides in Turn	ST	2.16 (0.37)	2.20 (0.41)
	DT	3.20 (0.41)	3.12 (0.53)

Results

Tables 2 and 3 report the ICC2, SEM (%SEM), and MDC_{95} for each gait parameter in ST-TUG and DT-TUG. In ST-TUG, very high reliability (ICC ≥ 0.90) was observed for time to complete the test, walk time, number of steps, mean stride time, cadence, and mean stride velocity. Good reliability (0.75 ≤ ICC < 0.90) was found for number of gait cycles, mean stance time, mean step time, mean stride length, return from turn time, and double support. Number of strides in turn, turn mid-point time, and turn time showed moderate reliability (0.5 ≤ ICC < 0.75). Coefficient of variability (CV) stride time, turning magnitude, CV stride velocity, CV stride length, CV step time, CV single support showed poor reliability (ICC < 0.5).

In DT-TUG, overall good reliability (0.75 ≤ ICC < 0.90) was observed for time to complete the test, walk time, number of steps, number of gait cycles, mean stride time, mean stride velocity, and mean stance time. Cadence, return from turn time, mean step time, mean stride time, double support, and turn time showed moderate reliability (0.50 ≤ ICC < 0.75). Turn mid-point time, number of strides in turn, CV step time, CV stride velocity, CV stride time, CV stride length, turning magnitude, and CV single support showed poor reliability (ICC < 0.50).

Across both tasks, SEM% for parameters with high and moderate confidence is minimal (SEM% < 7%), indicating small absolute measurement error, whereas CV-based and detailed turning measures had higher SEM%, limiting their usefulness for short-interval change detection. For high-reliability measure (complete the test, walk time, step time, step frequency, and step speed) were all less than 1 in the ST-TUG, and increased in the DT-TUG, but were mostly less than 1. The MDC_{95} for the coefficient of variability of gait and turning measures were higher in both conditions.

In summary, ST-TUG yielded higher ICC and lower SEM/MDC_{95} for most parameters than DT-TUG, reflecting more stable measurement. The DT-TUG exhibited moderate reliability on parameters such as time to complete the test, walk time, step stride, step length, and step speed, but its variability metrics and the turn metrics, exhibited substantial measurement errors and significantly diminished reliability. Clinically, timing/pace metrics (total/walk time, cadence/velocity, stride/step timing and length) emerge as priority candidates for monitoring

Table 3 Comparison of retest reliability of gait parameters on two occasions

Gait Parameters	Task Type	ICC (95%CI)	SEM (%SEM)	MDC_{95}
Time to Complete Test (s)	ST	0.938 (0.865–0.972)*	0.22 (2.48)	0.60
	DT	0.885 (0.756–0.948)	0.31 (2.68)	0.87
Walk Time (s)	ST	0.905 (0.796–0.957)*	0.19 (3.45)	0.51
	DT	0.899 (0.785–0.954)	0.23 (2.85)	0.62
Mean Stance Time (s)	ST	0.891 (0.769–0.951)	0.01 (1.32)	0.02
	DT	0.866 (0.720–0.939)	0.01 (0.98)	0.02
Mean Step Time (s)	ST	0.825 (0.643–0.919)	0.01 (2.19)	0.04
	DT	0.638 (0.331–0.822)	0.03 (4.80)	0.10
Mean Stride Time (s)	ST	0.927 (0.842–0.967)*	0.01 (1.24)	0.04
	DT	0.719 (0.458–0.865)	0.04 (2.93)	0.12
Mean Stride Length (cm)	ST	0.897 (0.780–0.953)	0.51 (0.75)	1.41
	DT	0.756 (0.520–0.884)	0.83 (1.12)	2.30
Mean Stride Velocity (cm/s)	ST	0.939 (0.867–0.973)*	0.40 (0.70)	1.10
	DT	0.810 (0.616–0.912)	0.47 (0.92)	1.29
Number of Gait Cycles	ST	0.831 (0.653–0.922)	0.23 (6.95)	0.65
	DT	0.758 (0.524–0.886)	0.32 (7.14)	0.89
Number of Steps	ST	0.909 (0.805–0.959)*	0.40 (3.91)	1.12
	DT	0.836 (0.663–0.924)	0.39 (2.82)	1.09
Cadence (steps/min)	ST	0.922 (0.831–0.965)*	1.39 (1.38)	3.84
	DT	0.668 (0.378–0.839)	1.44 (1.76)	3.99
Double Support % of gait cycle (%)	ST	0.796 (0.591–0.905)	0.62 (1.61)	1.72
	DT	0.675 (0.389–0.843)	1.42 (3.22)	3.94
Coefficient of Variability Stride Time (%)	ST	0.448 (0.072–0.712)	0.24 (8.61)	0.67
	DT	0.422 (0.040–0.696)	0.41 (9.47)	1.12
Coefficient of Variability Step Time (%)	ST	0.449 (0.073–0.713)	0.24 (7.35)	0.67
	DT	0.390 (0.002–0.676)	0.36 (7.62)	0.99
Coefficient of Variability Stride Velocity (%)	ST	0.436 (0.058–0.705)	0.36 (8.98)	1.00
	DT	0.448 (0.072–0.712)	0.36 (6.71)	0.99
Coefficient of Variability Stride Length (%)	ST	0.434 (0.056–0.704)	0.28 (9.83)	0.77
	DT	0.409 (0.025–0.688)	0.39 (10.05)	1.09
Coefficient of Variability Single Support (%)	ST	0.450 (0.075–0.714)	0.27 (8.06)	0.76
	DT	0.397 (0.010–0.680)	0.37 (7.47)	1.01
Turn Mid-Point Time (s)	ST	0.658 (0.363–0.834)	0.08 (3.94)	0.21
	DT	0.418 (0.035–0.694)	0.13 (4.78)	0.36
Turn Time (s)	ST	0.730 (0.476–0.871)	0.11 (4.81)	0.29
	DT	0.508 (0.149–0.749)	0.20 (6.82)	0.55
Return from Turn Time (s)	ST	0.753 (0.515–0.883)	0.05 (2.70)	0.13
	DT	0.701 (0.429–0.856)	0.08 (2.99)	0.21
Turning Magnitude (degrees/sec)	ST	0.478 (0.110–0.730)	4.12 (4.26)	11.42
	DT	0.424 (0.043–0.697)	2.08 (2.37)	5.77
Number Strides in Turn	ST	0.598 (0.273–0.800)	0.25 (11.36)	0.69
	DT	0.451 (0.076–0.714)	0.35 (11.11)	0.97

ICC Intraclass correlation coefficient, *SEM* Standard error of measurement, *SEM*% SEM expressed as a percentage of the mean, MDC_{95} Minimal detectable change at 95% confidence; An asterisk (*) indicates very high reliability (ICC ≥ 0.90)

and screening, whereas variability and turning measures are less suitable for detecting short-interval change.

Discussion

The study evaluated the retest reliability of healthy older women in ST-TUG and DT-TUG under wearable sensors. The present DT-TUG employed a cognitive–motor paradigm (serial subtraction by 3 s). Math-based cognitive load primarily engages working memory and sustained attention, which are executive resources that also support gait regulation, especially during turning and step-to-step control. Competition for these shared resources can create a central bottleneck, expressed as dual-task interference in motor and/or cognitive performance relative to single-task conditions. Within this framework, the reliability pattern observed here,

good-to-excellent reproducibility for core timing/pace metrics but lower short-interval reproducibility for variability and turn-related metric. Because different cognitive tasks emphasize partially distinct components, both the magnitude and locus of interference are task-dependent; accordingly, the reliability profile reported here should be interpreted in the context of a math-based cognitive load.

In the ST-TUG, the following parameters demonstrate good to very high test-retest reliability (ICC ≥ 0.75), including time to complete the test, walk time, cadence, number of steps, mean stride time, mean stride velocity, number of gait cycles, mean stance time, mean step time, return from turn time, mean stride length, and double support. SEM% for these variables was generally < 7%, indicating small absolute measurement error. These findings accord with prior work supporting the reliability of wearable sensor–based TUG assessments in older adults [8]. Wüest et al. examined stroke survivors and non-disabled older adults using an instrumented TUG and reported ICCs spanning poor to excellent (≈ 0.43–0.99), indicating heterogeneous reproducibility across metrics. Measurement error was small for several variables, underscoring the need to interpret reliability by metric rather than as a uniform property of the test [31]. Smith et al. measured the retest reliability of the ST-TUG and DT-TUG on five consecutive days using sensors, and the results showed that the return values for time to completion of the test, time to walk, number of gait cycles, number of steps, and turnaround time had moderate retest reliability ICC > 0.70 [26]. A meta-analysis further indicated that step counts and total time are more accurately quantified when inertial sensors are used alongside standardized timed walking tasks [18]. Rose et al. evaluated the reliability of wearable sensors for assessing walking and chair-standing activities at home in patients with knee osteoarthritis, and she found excellent retest reliability for core gait metrics (95% CI: 0.85–0.95) derived from wearable sensors [22].

In the DT-TUG, the time to complete the test, walk time, mean stance time, mean stride length, and mean stride velocity, number of gait cycles, and number of steps showed good agreement overall (0.75 ≤ ICC < 0.90), with SEM% mostly < 7.2%. Mean values for several timing/pace measures trended lower (i.e., slower or reduced pace) under dual-task load, but reliability remained acceptable. These observations are consistent with Coulthard et al., who reported good repeatability for DT-TUG timing measures using wearable technology [6]. Notably, although cognitive load can alter gait, prior work suggests that core timing/pace metrics may be relatively preserved, whereas effects on variability are task dependent [17].

By contrast, all the variability coefficients and turn-related parameters (e.g., turn midpoint time, turning magnitude) exhibited poor reliability (ICC < 0.50). This pattern aligns with Smith et al. in the quantitative TUG (QTUG), where gait variability indices and turn-related times showed weak retest reliability [26]. Using wearable sensors for QTUG, McGrath et al. reported good reliability for 18 variables but poor reliability for turn time [16]. Byun et al. used a triaxial accelerometer to assess the gait of older adults, thereby revealing an asymmetry in gait time variability and gait time. This asymmetry was found to have low retest reliability, requiring at least 26 and 14 steps, respectively, to achieve excellent reliability [4]. A review similarly reported poor-to-moderate validity and reliability for variability and symmetry measures [13]. Together with our data, where SEM% generally exceeded 6–10% and MDC_{95} thresholds were large for CV and turning metrics, these findings suggest that such measures are sensitive to individual variability, task complexity, and sample step counts, and are less suitable for detecting short-interval change than core spatiotemporal parameters.

Overall, in the presence of a cognitive load, spatiotemporal gait parameters retained good-to-moderate reliability and remained relatively stable, whereas CV and turn-related indices showed higher SEM% and larger MDC_{95}, indicating reduced reproducibility. This observation is indicative of an inadequate allocation of attentional resources, a factor that contributed to the instability of the sessions. This finding aligns with the "posture-first" strategy employed by older adults, which emphasises the maintenance of postural stability and fundamental gait patterns during the execution of concurrent cognitive and motor tasks [12, 24]. Practically, these results underscore the importance of selecting stable timing/pace metrics as primary endpoints for monitoring and screening in community settings using wearable technologies, while interpreting variability and turning measures with caution for short-interval change detection. These findings underscore the importance of selecting stable gait parameters when designing fall prevention protocols using wearable technologies in community settings.

Limitations and future directions

This study has several limitations. First, the sample comprised healthy, community-dwelling older women only; comparisons with older men and with clinical populations (e.g., neurological or musculoskeletal conditions) were not available, which limits generalisability. Future research should broaden the sample to include both sexes and diverse cognitive and pathological groups to enhance external validity. Second, we assessed short-interval test–retest reliability only (24–48 h). Multiple assessments

over longer intervals are needed to characterise longer-term reliability and variability of the metrics.

A further limitation is that cognitive-task accuracy during DT-TUG was not verified in real time. Although serial responses were recorded, correctness was not scored during testing; some participants may therefore have prioritised gait and produced incorrect or random answers. This could attenuate observable dual-task interference and bias reliability estimates, particularly for variability and turn-related metrics. Future protocols should implement real-time accuracy scoring with a priori thresholds and re-trial/exclusion rules, and compute both motor and cognitive dual-task costs to quantify trade-offs.

Theoretical and practical implications

This study quantified the test–retest reliability of TUG-derived gait parameters under both single-task and cognitive–motor dual-task conditions in older adults using wearable sensors. By extracting a multidimensional set of 21 parameters, including timing/pace, spatial, angular-velocity, and turn-related measures, we provide a more comprehensive characterisation of gait than studies focusing primarily on a small subset of variability or turning indices. The side-by-side comparison of ST-TUG and DT-TUG clarifies which parameters are stable over short intervals and which are more method-sensitive, thereby informing outcome selection in future studies and clinical monitoring. Although variability and turn-related measures showed limited short-interval reliability, DT-TUG, when combined with rigorous accuracy verification and detailed gait analysis, may add sensitivity for early risk screening relevant to falls and cognitive decline. These results provide data to guide metric selection, study design, and the implementation of sensor-based assessments in fall-prevention programs.

Conclusion

In community-dwelling older women, wearable-sensor–derived timed up and go metrics demonstrated high short-interval test–retest reliability for core spatiotemporal measures during ST-TUG, including total time, walking time, cadence, mean stride/step timing, mean stride velocity, stride length, number of steps and gait cycles, stance time, step time, return-from-turn time, and double support. By contrast, gait variability (CV) indices and turn-related measures (e.g., CV of stride time/velocity/length, CV of step time/single support, turning magnitude) showed poor reliability (ICC < 0.50). Under DT-TUG (cognitive–motor), the primary gait spatiotemporal parameters (time to complete the test, walk time, mean stance time, mean stride length//velocity, number of gait cycles/step) demonstrated good reliability overall. The gait variability coefficients and turn-related parameters (turn mid-point time, turning magnitude, CV stride time, CV stride velocity, CV stride length, CV step time, CV single support, number of strides in turn) retained overall good reliability, whereas variability and turn-related parameters again showed lower reproducibility.

The findings provide substantial evidence that ST-TUG and DT-TUG are reliable tools for assessing fall risk in older adults. In this paper, we propose a method for the more intuitive assessment and screening of fall risk in older adults, utilising quantitative analysis of various gait parameter metrics. This approach offers a reliable alternative to traditional tools, such as questionnaire-based methods, which can be challenging to quantify.

From a practical standpoint, accelerometer-based assessment enables scalable, objective capture of gait in everyday environments, facilitating longitudinal follow-up and potential early-warning signals relevant to fall-risk management across community screening, telemedicine, and home-based rehabilitation. Future work should validate cognitive-task accuracy during DT-TUG, compare multiple dual-task paradigms head-to-head, and evaluate longer retest intervals and real-world monitoring protocols to strengthen generalisability and optimise metric selection for fall-prevention programs.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12877-025-06870-0.

Supplementary Material 1.

Supplementary Material 2.

Authors' contributions

Xiao Yu: Conceptualization, Methodology, Data Collection, Formal Analysis, Visualization, Writing – Original Draft. Roxana Dev Omar Dev: Supervision, Writing – Review & Editing, Validation, Project Administration.Both authors have read and approved the final version of the manuscript.

Funding

No funding was received.

Data availability

The raw data in this study are available from the corresponding author (Xiao Yu and Roxana Dev Omar Dev). All data were collected under ethical approval (Universiti Putra Malaysia Ethics Committee, Approval No. JKEUPM-2023-373) and anonymized to protect the privacy of the participants.

Declarations

Ethics statement and informed consent

This study was approved by the Ethics Committee of Universiti Putra Malaysia (Approval No. JKEUPM-2023-373). All procedures involving human participants were conducted in accordance with the Declaration of Helsinki and its subsequent amendments. Written informed consent was obtained from all participants prior to study enrollment.

Consent for publication

Not Applicable.

Competing interests

The authors declare no competing interests.

Received: 18 May 2025 / Accepted: 27 November 2025
Published online: 15 December 2025

References

1. Atkinson C, Earnshaw P. Motivational cognitive behavioural therapy: distinctive features. 1st ed. London: Routledge; 2019. p. 188.
2. Barry E, Galvin R, Keogh C, Horgan F, Fahey T. Is the timed up and go test a useful predictor of risk of falls in community dwelling older adults: A systematic review and meta-analysis. BMC Geriatr. 2014;14:1–14.
3. Beauchet O, Fantino B, Allali G, Muir S, Montero-Odasso M, Annweiler C. Timed up and go test and risk of falls in older adults: A systematic review. J Nutr Health Aging. 2011;15:933–8.
4. Byun S, Han JW, Kim TH, Kim KW. Test-retest reliability and concurrent validity of a single tri-axial accelerometer-based gait analysis in older adults with normal cognition. PLoS ONE. 2016;11(7):e0158956.
5. Choi J, Parker SM, Knarr BA, Gwon Y, Youn J-H. Wearable sensor-based prediction model of timed up and go test in older adults. Sensors. 2021;21(20):6831.
6. Coulthard JT, Treen TT, Oates AR, Lanovaz JL. Evaluation of an inertial sensor system for analysis of timed-up-and-go under dual-task demands. Gait Posture. 2015;41(4):882–7.
7. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: A guide for non-statisticians. Int J Endocrinol Metabolism. 2012;10(2):486.
8. Greene BR, O'Donovan A, Romero-Ortuno R, Cogan L, Scanaill CN, Kenny RA. Quantitative falls risk assessment using the timed up and go test. IEEE Trans Biomed Eng. 2010;57(12):2918–26.
9. Guan Q, Han H-J, Zhan Q, Yun Y-X, Zhao L-J. Activities-specific balance confidence (ABC) scale adapted for the Mainland population of China. Clin Rehabil. 2012;26(7):648–55.
10. Guan Q, Han H-J, Zhan Q, Yun Y-X, Zhao L-J. Research of reliability and validity of Chinese translated Activities-Specific balance confidence scale. J Tongji Univ (Medical Science). 2011;32(03):81–4.
11. Harvill LM. Standard error of measurement: an NCME instructional module on. Educational Measurement: Issues Pract. 1991;10(2):33–41.
12. Huxhold O, Li S-C, Schmiedek F, Lindenberger U. Dual-tasking postural control: aging and the effects of cognitive demand in conjunction with focus of attention. Brain Res Bull. 2006;69(3):294–305.
13. Kobsar D, Charlton JM, Tse CT, Esculier J-F, Graffos A, Krowchuk NM, Thatcher D, Hunt MA. Validity and reliability of wearable inertial sensors in healthy adult walking: A systematic review and meta-analysis. J Neuroeng Rehabil. 2020;17:1–21.
14. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63.
15. Kressig RW, Herrmann FR, Grandjean R, Michel J-P, Beauchet O. Gait variability while dual-tasking: fall predictor in older inpatients? Aging Clin Exp Res. 2008;20:123–30.
16. McGrath D, Greene BR, Doheny EP, McKeown DJ, De Vito G, Caulfield B. Reliability of quantitative TUG measures of mobility for use in falls risk assessment. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2011. p. 466–9.
17. Montero-Odasso M, Casas A, Hansen KT, Bilski P, Gutmanis I, Wells JL, Borrie MJ. Quantitative gait analysis under dual-task in older people with mild cognitive impairment: A reliability study. J Neuroeng Rehabil. 2009;6:1–6.
18. Montesinos L, Castaldo R, Pecchia L. Wearable inertial sensors for fall risk assessment and prediction in older adults: A systematic review and meta-analysis. IEEE Trans Neural Syst Rehabil Eng. 2018;26(3):573–82.
19. Nordin E, Moe-Nilssen R, Ramnemark A, Lundin-Olsson L. Changes in step-width during dual-task walking predicts falls. Gait Posture. 2010;32(1):92–7.
20. Podsiadlo D, Richardson S. The timed up & go: A test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39(2):142–8.
21. Rose DJ, Jones CJ, Lucchese N. Predicting the probability of falls in community-residing older adults using the 8-foot up-and-go: A new measure of functional mobility. J Aging Phys Act. 2002;10(4):466–75.
22. Rose MJ, Neogi T, Friscia B, Torabian KA, LaValley MP, Gheller M, Adamowicz L, Georgiev P, Viktrup L, Demanuele C. Reliability of wearable sensors for assessing gait and chair stand function at home in people with knee osteoarthritis. Arthritis Care Res. 2023;75(9):1939–48.
23. Schoene D, Valenzuela T, Lord SR, de Bruin ED. The effect of interactive cognitive-motor training in reducing fall risk in older people: A systematic review. BMC Geriatr. 2014;14:1–22.
24. Shumway-Cook A, Woollacott M, Kerns KA, Baldwin M. The effects of two types of cognitive tasks on postural stability in older adults with and without a history of falls. Journals Gerontol Ser A: Biol Sci Med Sci. 1997;52(4):M232–40.
25. Siu K-C, Chou L-S, Mayr U, van Donkelaar P, Woollacott MH. Does inability to allocate attention contribute to balance constraints during gait in older adults? Journals Gerontol Ser A: Biol Sci Med Sci. 2008;63(12):1364–9.
26. Smith E, Walsh L, Doyle J, Greene B, Blake C. The reliability of the quantitative timed up and go test (QTUG) measured over five consecutive days under single and dual-task conditions in community dwelling older adults. Gait Posture. 2016;43:239–44.
27. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris back pain questionnaire: part 1. Phys Ther. 1998;78(11):1186–96.
28. Stratford PW, Riddle DL. When minimal detectable change exceeds a diagnostic test–based threshold change value for an outcome measure: resolving the conflict. Phys Ther. 2012;92(10):1338–47.
29. Tong Y, Rong J, Tian X, Wang Y, Chen Z, Adams R, et al. Use of dual-task timed-up-and-go tests for predicting falls in physically active, community-dwelling older adults—a prospective study. J Aging Phys Act. 2023;31(6):948–55.
30. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–40.
31. Wüest S, Massé F, Aminian K, Gonzenbach R, De Bruin ED. Reliability and validity of the inertial sensor-based timed up and go test in individuals affected by stroke. J Rehabilitation Res Dev. 2016;53(5):599–610.
32. World Health Organization. (2021). Falls. https://www.who.int/news-room/fact-sheets/detail/falls.
33. Wyrwich KW. Minimal important difference thresholds and the standard error of measurement: is there a connection? J Biopharm Stat. 2004;14(1):97–110.
34. Zasadzka E, Borowicz AM, Roszak M, Pawlaczyk M. Assessment of the risk of falling with the use of timed up and go test in the elderly with lower extremity osteoarthritis. Clin Interv Aging. 2015;10:1289–98.

Publisher's Note