

Original Paper

A Smartphone-Based Timed Up and Go Test Self-Assessment for Older Adults: Validity and Reliability Study

Melissa Johanna Böttinger^{1,2}, MSc; Sabato Mellone³, PhD; Jochen Klenk^{4,5}, Prof Dr; Carl-Philipp Jansen¹, PhD; Marios Stefanakis⁶, MSc; Elena Litz^{1,2}, MSc; Anastasia Bredenbrock², BSc; Jan-Philipp Fischer⁷, MA; Jürgen M Bauer^{1,2}, MD; Clemens Becker¹, MD; Katharina Gordt-Oesterwind^{1,2,8}, PhD

¹Geriatric Center, Medical Faculty Heidelberg, Heidelberg University, Heidelberg, Germany

²Network Aging Research, Heidelberg University, Heidelberg, Germany

³Department of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy

⁴Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany

⁵Study Center Stuttgart, IB University of Health and Social Sciences, Stuttgart, Germany

⁶Department of Clinical Gerontology and Geriatric Rehabilitation, Robert Bosch Hospital, Stuttgart, Germany

⁷TSG ResearchLab gGmbH, Zuzenhausen, Germany

⁸Institute of Sports and Sports Sciences, Heidelberg University, Heidelberg, Germany

Corresponding Author:

Melissa Johanna Böttinger, MSc
Geriatric Center, Medical Faculty Heidelberg
Heidelberg University
Bergheimer Str. 20
Heidelberg, 69115
Germany
Phone: 49 6221 548146
Email: melissajohannaboettinger@gmail.com

Abstract

Background: The Timed Up and Go test (TUG) is recommended as an evidence-based tool for measuring physical capacity. Instrumented TUG (iTUG) approaches expand classical supervised clinical applications offering the potential of self-assessment for older adults.

Objective: This study aimed to evaluate the concurrent validity and test-retest reliability of a smartphone-based TUG self-assessment “up&go app.”

Methods: A total of 52 community-dwelling older adults (>67 years old) were recruited. A validated and medically certified system attached with a belt at the lower back was used as a reference system to validate the “up&go app” algorithm. The participants repeated the TUG 5 times wearing, a smartphone with the “up&go app” in their front trouser pocket and an inertial sensor to test the concurrent validity. A subsample of 37 participants repeated the “up&go app” measurement 2 weeks later to examine the test-retest reliability.

Results: The correlation between the “up&go app” and the reference measurement was $r=0.99$ for the total test duration and $r=0.97$ for the 5 single repetitions. Agreement between the 5 repetitions was intraclass correlation coefficient (ICC)=0.9 (0.84-0.94). Leaving out the first repetition, the agreement was ICC=0.95 (0.92-0.97). Test-retest agreement had an ICC=0.79 (0.53-0.9).

Conclusions: The duration of 5 repetitions of the TUG test, measured with the pocket-worn “up&go app,” was very consistent with the results of a lower-back sensor system, indicating excellent concurrent validity. Participants walked slower in the first round than in the other 4 repetitions within a test run. Test-retest reliability was also excellent. The “up&go app” provides a useful smartphone-based approach to measure 5 repetitions of the TUG. The app could be used by older adults as a self-screening and monitoring tool of physical capacity at home and thereby help to early identify functional limitations and take interventions when necessary.

JMIR Aging 2025;8:e67322; doi: [10.2196/67322](https://doi.org/10.2196/67322)

Keywords: timed up and go test ; self-assessment; instrumented assessment; technology-based assessment; physical capacity; mobility; aged; mobile applications; smartphone; diagnostic self evaluation

Introduction

The baby boomer generation embraces smartphone technology [1]. This particular cohort is willing and capable of using mobile health applications especially when they are designed in a user-centered way and increase the likelihood of a self-determined future [2-4]. Good examples are the use of fitness trackers, heart rate monitoring, and glucose monitoring [5].

The World Health Organization (WHO) and United Nations (UN) member states have declared the current decade as the UN “Decade of Healthy Ageing” (2021-2030) to improve the lives of older people, their families, and the communities in which they live [6-8]. Early identification of health risks and functional limitations in older adults is crucial to enable timely preventive measures and promote the maintenance of independence [9]. A recent systematic review was conducted to guide WHO recommendations and to establish a “standard” for evidence-based assessment of physical capacity [10]. It recommends the Timed Up and Go Test (TUG), in addition to the 30-Second Chair Rise Test, due to the high quality evidence for its sufficiently valid and reliable measurement of physical capacity in community-dwelling older adults [10,11]. The standard TUG captures the time recorded by a trained test administrator using a stopwatch. The test person rises from a chair, walks 3 meters at their usual pace, turns around, walks back, and sits down again. Longer test durations indicate limitations in physical capacity, particularly in strength and balance [10-12] and predict hospitalization and associated functional decline in older adults [13]. The need for and lack of a digital version of the TUG has been addressed during the meetings of the WHO Locomotor Capacity Working Group. This is even more relevant for persons living remotely or in low and middle-income countries where smartphone technology is often available but health care professionals are often lacking [2].

Wearable technologies offer opportunities to lower the barrier for integrating functional tests like the TUG into the assessment and clinical management of older people [14,15]. Such an instrumented Timed Up and Go Test (iTUG) not only enables standardized and digitized measurement in supervised settings [16-18] but also opens an option for low-cost self-screening and monitoring in people’s everyday lives [19,20]. In clinical studies, camera-based systems, inertial sensors or smartphones that are attached to the back (eg, with a belt) are most commonly used to perform the iTUG [21]. For self-assessment purposes, however, an approach using common technologies and a convenient placement on the body would be more feasible, without the need for additional materials or training. Therefore, the “up&go” iTUG was developed as a self-test for older adults. The prototype of the “up&go” app emerged from several EU projects [17-19,22,23] and a cocreation process with older adults [24]. It incorporates an innovative approach: placing a smartphone

in the user’s front trouser pocket during 5 repetitions of the test. An unsupervised smartphone-based iTUG could fill a gap in the landscape of mobility assessment methods that is still dominated by patient-reported outcome measures and supervised assessments conducted in laboratory or clinical settings [25]. Digital self-assessments could support older adults to take up an active role in risk screening or monitoring, and to early identify changes in their mobility. If necessary, this would allow them to initiate further medical diagnostics, and to take primary or secondary preventive measures in a timely manner. Thus, the motivation for this study is to examine if the “up&go” self-assessment app fulfils the criteria of validity and reliability to be considered trustworthy and to serve as a credible tool for implementation in future iTUG studies and in the self-management of older people.

Hence, this study aims to examine the concurrent validity and test-retest reliability of the “up&go” smartphone-based TUG self-assessment for older adults. The measures considered are the total TUG test duration and the duration of the 5 individual repetitions. This study corresponds to an analytical validation according to the V3 framework [26].

The objectives of this study are to analyze the concurrent validity of the algorithm used in the “up&go app” against a validated and medically certified sensor-based system and to determine the test-retest reliability of the “up&go app” algorithm in the home environment of older adults.

Methods

Participant Recruitment

The findings are reported following the guidelines for reporting reliability and agreement studies [27]. Participant were recruited from the SMART-AGE randomized controlled trial conducted at Heidelberg University, Germany [28]. Participants were included in the SMART-AGE study if they were 67 years or older, lived in the community, had basic knowledge of using PCs or tablets and of the German language. Exclusion criteria were severe medical conditions (ie, heart failure with shortness of breath at rest, cardiac arrhythmia with dizziness, Parkinson disease with use of a walker or wheelchair, cancer with chemotherapy or radiotherapy, chronic lung disease with oxygen therapy, and planned major medical procedure with inpatient hospitalization within the next 3 months), severe visual or hearing impairment, and severe cognitive impairment.

Data for the sample description were extracted from the SMART-AGE initial assessment data: age, sex, BMI, citizenship, living situation, employment, WHO Quality of Life Scale score [29], Satisfaction With Life Scale score [30], fall history, Groll Functional Comorbidity Index [31], Trail Making Test A and B score [32], Six-item Cognitive Impairment Test score [33], Short Falls Efficacy Scale International score [34], 4-Meter Walk Test score [35],

30-second Chair Rise Test score [36], and stopwatch-measured Timed-Up and Go Test score [11].

Ethical Considerations

The SMART-AGE study and the amended protocol for this sub study were approved by Heidelberg University Medical Faculty's Ethical Committee (S-672/2022). All participants provided informed signed consent before participating.

Data Collection and Processing

The data collection was performed during a home visit 3 to 4 months after the SMART-AGE study's baseline assessment. A total of 3 pretrained assessors with a professional background in psychology (AB and EL) or physiotherapy (MJB) conducted the home visits following a standardized manual.

“Up&go App”: Smartphone-Based Timed Up and Go Test

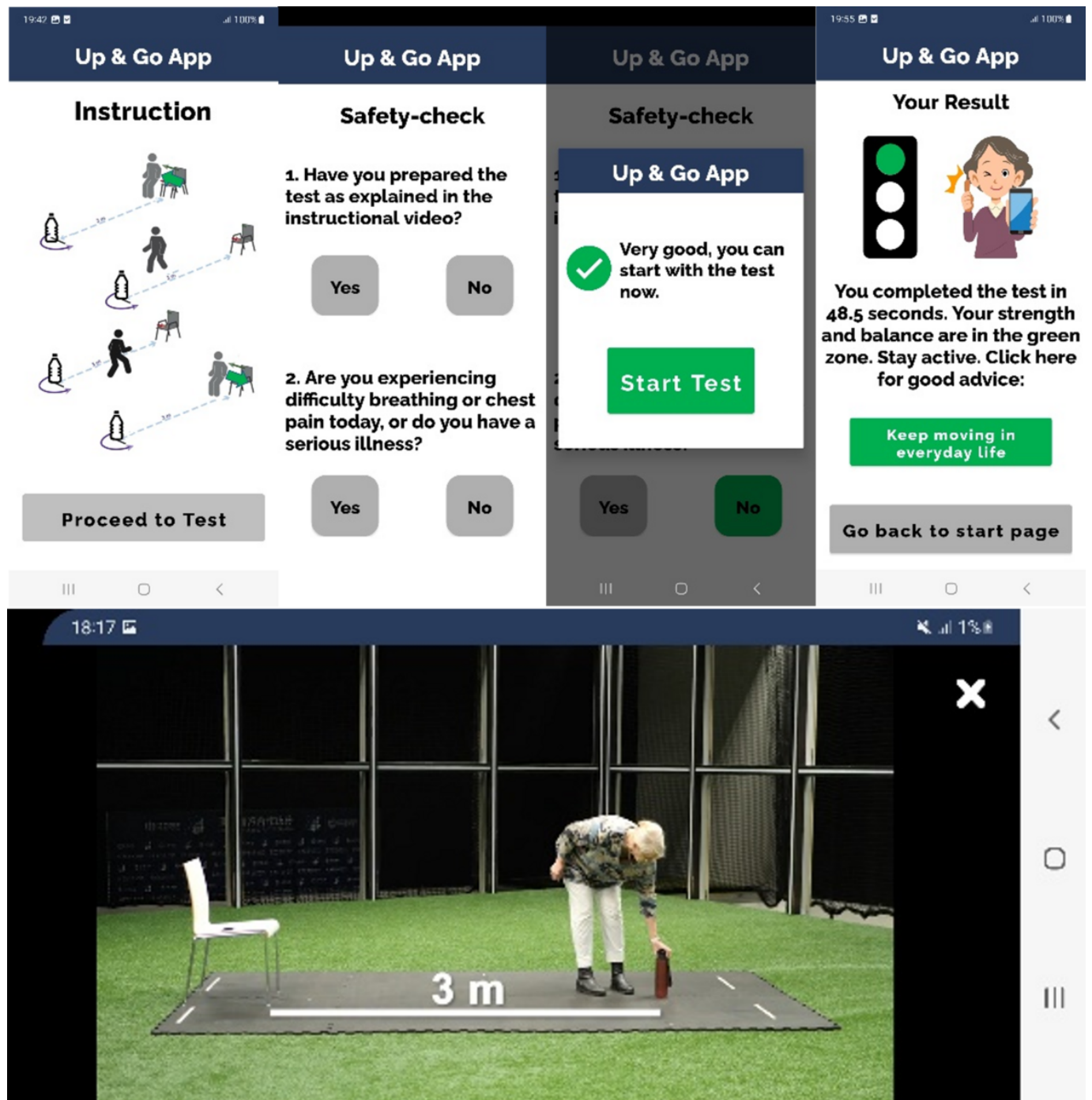
Previous versions of the “up&go app” algorithm have been developed and validated within the 2 European Commission funded projects, Farseeing [17,22] and PreventIT [18,19,23]. The “up&go app” investigated in this study [37] is a further development of the previous versions. It is based on the results of a study examining the usability of the previous Norwegian version of the PreventIT project [19] and a cocreation study with older adults [24].

In order to run the TUG test within the “up&go app,” the user is invited to watch and read the test instructions, prepare

the test setting (chair, 3 m distance, marker), answer 2 safety questions, press the “start test” button, place the smartphone in one of the front trouser pockets and sit down to begin with the first round of the test at usual gait speed (Figure 1). The app includes an algorithm processing real-time accelerometer data to detect the body movements to guide the user through 5 consecutive repetitions of the TUG via audio prompts (eg, “Please put the phone in your front trouser pocket and sit down”). After each run there is a pause of about 3 seconds in which the algorithm detects the position change, and the new audio announcement is played (eg, “This was the third run. Please wait for the next start signal”). The embedded sensors in the smartphone collect data during the test.

After the test, the algorithm automatically postprocesses accelerometer data and computes the total time needed to complete 5 repetitions, excluding pauses between runs. The total time is shown on the smartphone screen in seconds together with a traffic light (Figure 1, green= ≤ 60 s, yellow 60-90 s, red >90 s) and an according recommendation for action (eg, green light: recommendation to stay active, yellow light: recommendation for supervised training, and red light: recommendation to seek medical advice). The color coding thresholds are based on the cut-off value of 12 seconds for 1 repetition, respectively 60 seconds for 5 repetitions, to indicate normal versus below normal mobility in community-dwelling older adults [13,38].

Figure 1. Screenshots from the “up&go app”.

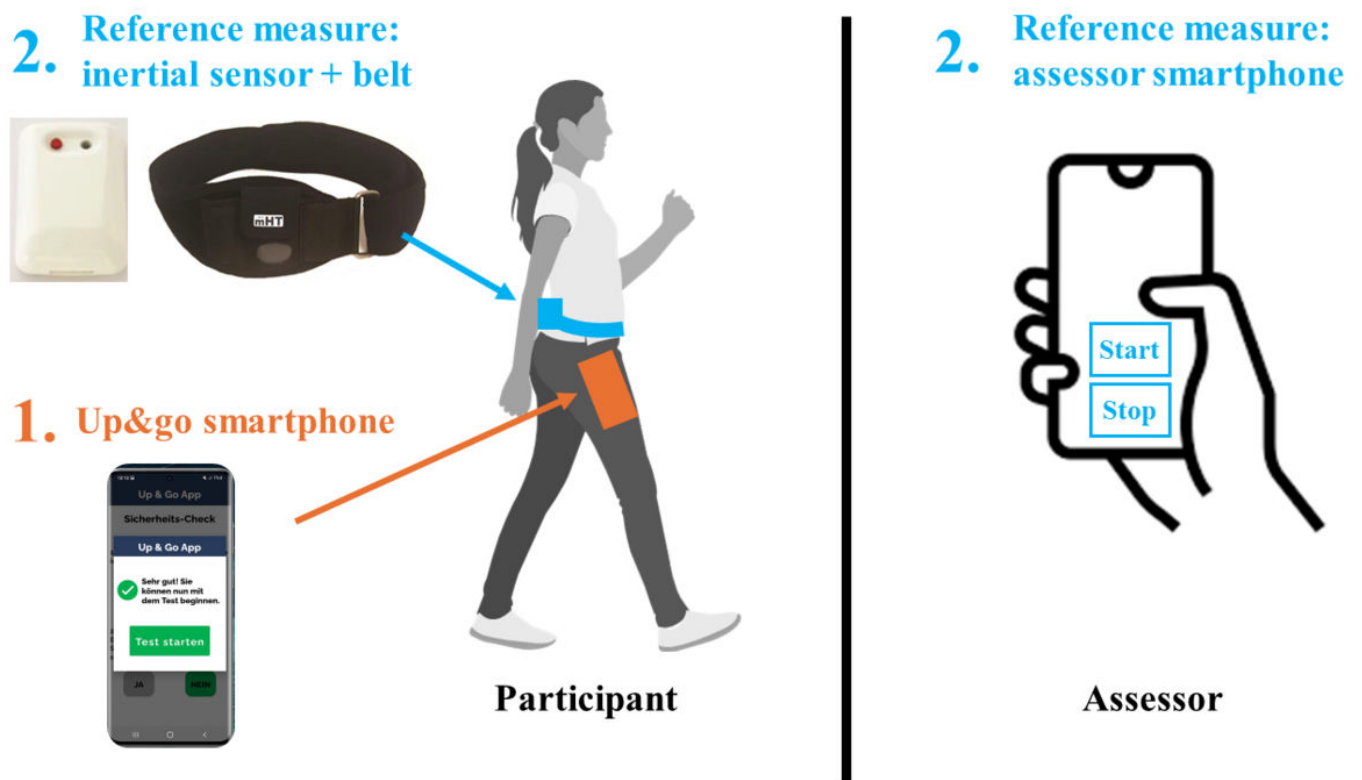


Concurrent Validity

In this study, the “mTUG” system (“mTUG” medical device, mHealth Technologies) was used as a reference measure to validate the “up&go app” algorithm. “mTUG” is a sensor-based system that has been validated in a population of older adults [17] and is certified as a medical device. The system consists of an inertial sensor (mHealth Technologies triaxial accelerometer range ± 2 g, triaxial gyroscope range $\pm 250^\circ/\text{sec}$, sampling rate: 100 Hz) connected by Bluetooth to an assessor smartphone (Samsung Galaxy S10e, Android 5.0.1) including a customized app (mHealth Technologies) allowing the assessor to manually start and stop the sensor measurement.

The TUG was repeated 5 times with participants wearing the 2 measurement systems simultaneously (Figure 2), smartphone with “up&go app” in their right front trouser pocket and an inertial sensor (mHealth Technologies) attached to an elastic belt worn on the lower back. A previous study with hospitalized hip fracture patients found that the performance of three TUG trials is recommended to achieve performance stability [39]. In this study, 5 repetitions were performed to take into account more variability in the performance of older adults conducting the test at home.

Figure 2. : Simultaneous measurement with (1) the “up&go app” and (2) the reference measure.



Participants were familiar with the TUG procedure as the conventional stopwatch TUG [11] had been performed beforehand as part of the baseline motor assessments within the SMART-AGE study. As the feasibility of the “up&go app” self-assessment was not part of this study, test preparation, viewing of instructional videos by participants were omitted for data collection. The test procedure was explained to the study partner (5 repetitions, usual walking pace, and follow audio instructions). The assessor prepared the TUG test setting, mounted the sensor in the belt on the participant’s lower back and started the test on the “mTUG” assessor smartphone. Then, the assessor started the test on the “up&go” smartphone and handed it over to the participant. The participant then placed it in the right front trouser pocket and performed the 5 repetitions of the TUG test, following the audio instructions of the “up&go app.”

The “up&go app” does not store any data on the user’s smartphone to guarantee data confidentiality. We therefore developed an adapted version for this study, allowing the app to store all raw data collected from the accelerometer and gyroscope embedded in a Samsung Galaxy S21 smartphone (Android 5.0.1, triaxial accelerometer range ± 8 g, triaxial gyroscope range $\pm 1000^\circ/\text{sec}$, sampling rate 500 Hz) on the smartphone memory. This allowed postprocessing of the raw data and analysis of not only the total test duration, but also the duration of each of the 5 repetitions (rep1, rep2, rep3, rep4, and rep5).

Within the “mTUG” system, raw data was automatically stored on the assessor smartphone’s memory and processed by a validated algorithm [17].

Test-Retest Reliability

To investigate the test-retest reliability of the “up&go app” algorithm, we repeated the “up&go” measurement at an additional home visit within 2 weeks after the first measurement, following the same procedure as described above.

Participants were asked to participate in the retest if they lived <20 km from the study center to reduce the burden of assessor time.

Statistical Analysis

The sample size for the validation part of the study was calculated using R (version 2023.09.1, R Foundation for Statistical Computing, package “pwr”). A strong correlation of ≥ 0.75 was assumed. With a power of 95% and a statistical significance level of $\alpha = .05$, ≥ 45 participants are required. The sample size for the test-retest part of the study was calculated using R (version 2023.09.1, package “ICC.sample.size”). An intraclass correlation coefficient (ICC) of ≥ 0.75 was assumed. With power=95% and $\alpha = .05$, ≥ 37 participants were required for 2 measurements (first measurement vs retest). Assuming a 15% dropout rate, 52 participants were recruited for the study. From this sample, we consecutively recruited participants for the retest measurement until a subsample size of at least 37 participants was reached to examine the test-retest reliability of the “up&go app” algorithm. The distribution of the data was checked before analyses (Shapiro-Wilk test). All data fulfilled the normal distribution assumption. For all statistical tests, $\alpha = .05$ was used as the threshold for statistical significance. Pearson correlation coefficients were calculated to examine the association between the systems (“up&go

app” and reference measure) regarding the total duration needed to complete the test as well as the duration of the 5 single repetitions. Correlation coefficients of 0-0.19 are interpreted as very weak, 0.2-0.39 as weak, 0.4-0.59 as moderate, 0.6-0.79 as strong, and 0.8-1 as very strong [40]. For graphical description of the agreement between the 2 systems (“up&go app” and reference measure), Bland-Altman plots were used [41], including the lower and upper limits of agreement (LLOA and ULOA). From the data collected during the first measurement (“up&go app” and reference measure), ICC 3,1 were calculated to analyze the agreement between the 5 repetitions. Test-retest reliability of the “up&go app” was evaluated using ICC [1,2,42]. An ICC value <0.4 indicates poor reliability, an ICC value between ≥ 0.4 and <0.75 fair to good reliability and an ICC value ≥ 0.75 indicates excellent reliability [43]. It was assumed that the association between “up&go app” and the reference measure is very strong ($r > 0.8$) and the agreement between “up&go” and the reference measure is good (Bland-Altman plot). Furthermore, we assumed that the test-retest reliability of the “up&go” test meets excellent levels (ICC ≥ 0.75). Statistical analyses were computed using statistical software R (version 2023.09.1) and MATLAB (version R2022b, MathWorks).

Results

Participant Data

A total of 52 community-dwelling older adults aged between 66 and 88 years (mean 73.6, SD 5.4) were measured, with a similar distribution of men (25/52, 48%) and women (27/52, 52%) and a mean BMI of 27.1 kg/m² (SD 4.1). Of those who provided the data, all were retired, almost all were German citizens (49/52, 98%) and one third (16/52, 33%) lived alone. Participants indicated an average of 2.3 (SD 1.7) comorbidities according to the Groll Index. Of these, degenerative disc disease (17/52), visual impairment (13/52), depression (11/52), and osteoarthritis (11/52) were reported most frequently. Cognition was normal (<7 errors on the Six-item Cognitive Impairment Test) and executive functions (Trail Making Test A and B) were in accordance with normative data of this age group [44]. The average duration of the TUG measured with the stopwatch in the SMART-AGE initial assessment was 9.7 (SD 1.7) seconds. None of the study participants used a walking aid during the TUG. Table 1 provides a detailed description of the sample.

Table 1. Characteristics of the study sample.

Data from the SMART-AGE initial assessment ^a	Values	Available sample
Age (years), mean (SD)	73.6 (5.4)	52
Female sex, n (%)	27 (52)	52
BMI (kg/m ²), mean (SD)	27.1 (4.1)	52
German citizenship, n (%)	49 (98)	50
Living alone, n (%)	16 (33)	49
Retired, n (%)	50 (100)	50
Stopwatch-measured TUG (s), mean (SD)	9.7 (1.7)	52
4-meter walk test, mean (SD)		
Duration (s)	4.2 (0.9)	52
Gait speed (m/s)	1.0 (0.2)	52
30-Second chair rise test, mean (SD)		
Number of repetitions	12.3 (3.3)	52
Short Falls Efficacy Scale International, mean (SD)		
Score	8.2 (2.2)	49
Groll Functional Comorbidity Index (score), mean (SD)	2.3 (1.7)	48
Fall history (“Yes, I fell in the last 12 months”) ^b , n (%)	19 (37)	51
Trail making test, mean (SD)		
Test duration A (s),	45.2 (16.6)	52
Test duration B (s)	99 (50.3)	52
Ratio (B/A)	2.3 (1.1)	52
Six-item cognitive impairment test^c, mean (SD)		
Score	1.0 (1.5)	52
Satisfaction with life scale, mean (SD)		
Score	5.0 (1.2)	49
WHO Quality of Life Scale, mean (SD)		
Physical domain (score)	77.3 (13.1)	49
Psychological domain (score)	72.4 (10.3)	49

Data from the SMART-AGE initial assessment ^a	Values	Available sample
Social relationships domain (score)	65.3 (13.6)	49
Environment domain (score)	79.8 (10.1)	49

^aThe initial SMART-AGE assessment took place 3-4 months before the data collection for this study.

^bFall history data was obtained one month after the initial SMART-AGE assessment.

^cParticipants with >7 error points were excluded from the SMART-AGE study due to a suspected dementia disorder.

Concurrent Validity

In total, 35 (67.3%) participants completed the test successfully on the first attempt. Due to technical problems (eg, smartphone shifted in trouser pocket), 12 (23.1%) participants had to repeat the test a second time, 4 (7.7%) participants a third time, and 1 (1.9%) participant a fourth time. After an unsuccessful attempt, participants always took a break of at least 5 minutes. As the app only saves data from complete test runs, failed attempts were not included.

On average, participants needed 52.4 (SD 7.8) seconds to complete 5 repetitions of the TUG, as measured by the

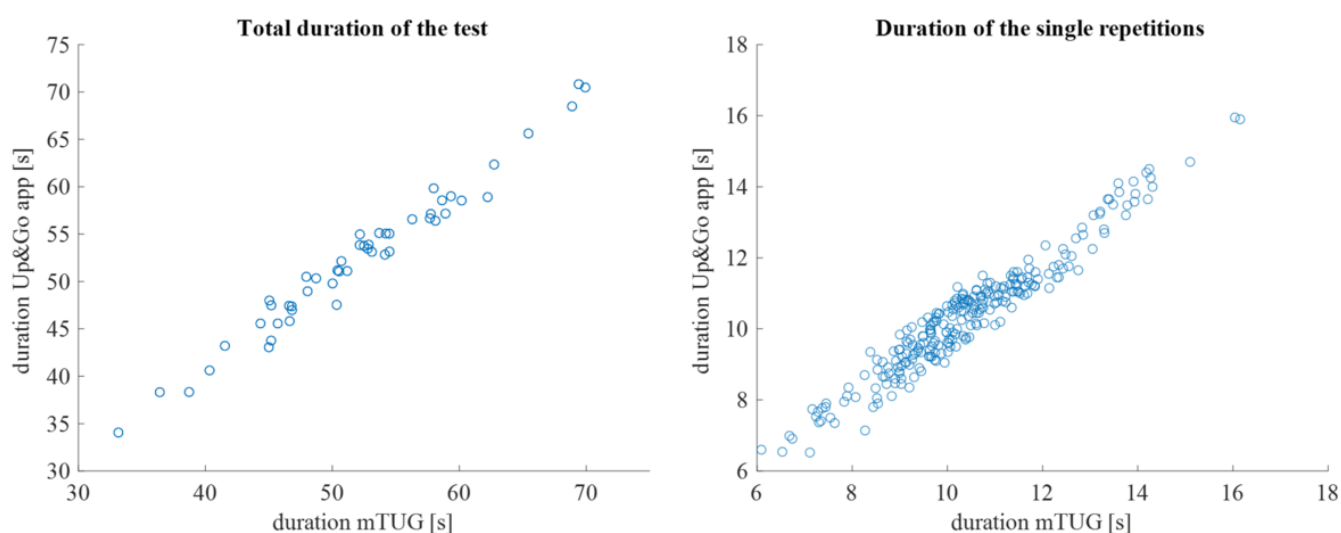
“up&go app”. Table 2 shows the results for the total duration of the test and the duration of the 5 repetitions as measured by the “up&go app” and the reference measure.

Data from all 52 participants were available for the concurrent validity analysis (“up&go app” vs reference measure). The Pearson correlation coefficient (r) for the total duration was 0.99 (95% CI 0.98-0.99) and that for the durations of the 5 single repetitions was 0.97 (95% CI 0.96-.97; Figure 3).

Table 2. Total test duration and duration of the 5 repetitions measured by the “up&go app” and reference measure.

Duration	“up&go app”, mean (SD)	Reference measure, mean (SD)
Total duration of the test (s)	52.38 (7.8)	52.31 (8.11)
First repetition (s)	10.99 (2.1)	11.11 (2.17)
Second repetition (s)	10.33 (1.54)	10.39 (1.61)
Third repetition (s)	10.25 (1.57)	10.34 (1.64)
Fourth repetition (s)	10.21 (1.52)	10.2 (1.52)
Fifth repetition (s)	10.11 (1.46)	10.26 (1.6)

Figure 3. Correlation analysis between “up&go app” and reference measure: total duration and duration of the single repetitions. TUG: Timed Up and Go Test.



The agreement between the total duration measured by the “up&go app” and the reference measure is shown in Figure 4 (Bland-Altman plots). According to Bland-Altman analysis, the mean difference was -0.27 for the total duration of the test and -0.05 for the duration of the single repetitions. The difference in the total duration of the test was between

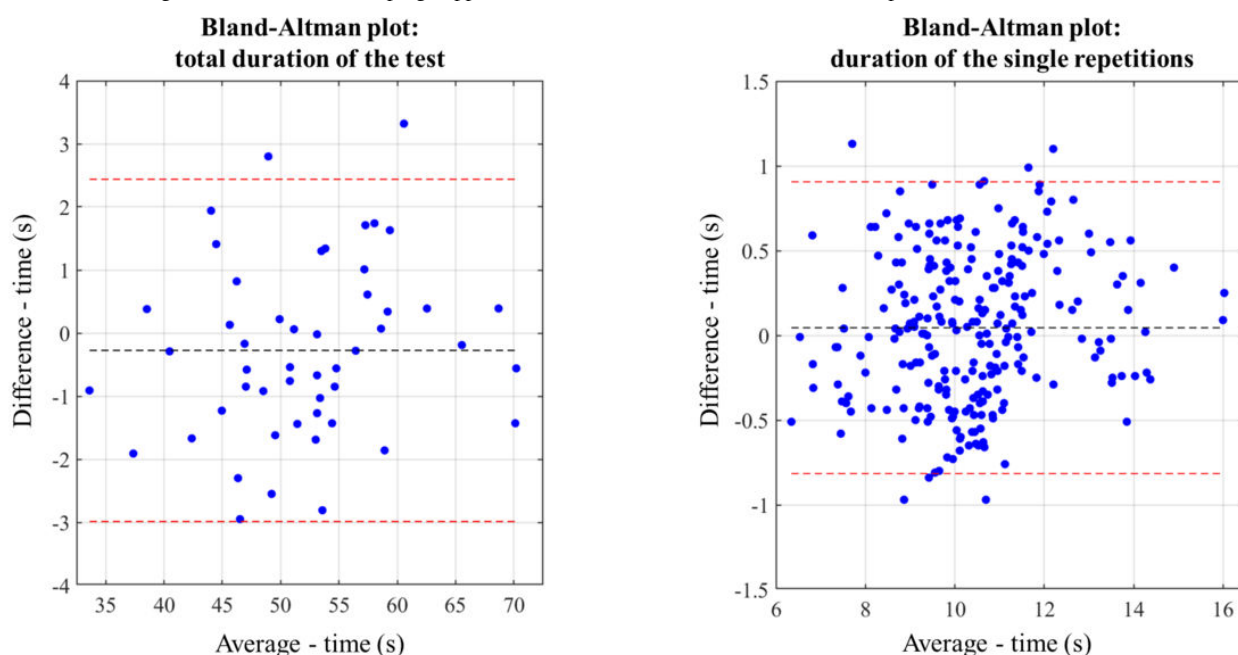
-2.99 seconds (LLoA) and -2.44 seconds (ULoA) and the difference in the duration of a single repetition was within -0.82 seconds (LLoA) and -0.91 seconds (ULoA) with respect to the estimate of the reference measure.

Agreement between all “up&go app” repetitions was $ICC(3,1)=0.9$ (95% CI 0.84-0.94) and $ICC(3,1)=0.87$ (95%

CI 0.80-0.93) between all repetitions measured by the reference measure. Leaving out the first repetition, the agreement of repetitions 2-5 was ICC(3,1)=0.95 (95% CI

0.92-0.97) for the “up&go app” and ICC(3,1)=0.93 (95% CI 0.89-0.96) for the reference measure.

Figure 4. Statistical agreement between the “up&go app” and the reference measure—Bland-Altman plot.



Test-Retest Reliability

Of the 52 participants, 37 were measured twice, and 1 participant was excluded due to noncompliance with the retest instruction to walk at habitual gait speed. Retests were performed on average 6 (SD 2.6) days after the first measurement. At the “up&go app” retest measurement, participants needed on average 48.6 (SD 8.5) seconds to complete the 5 repetitions. The ICC between the total duration during the first home visit and during the retest was ICC(2,1)=0.79 (95% CI 0.53-0.9).

Discussion

Principal Findings

To enable self-assessment for older people in their own homes, the “up&go app” was developed as a pocket-worn approach with 5 repetitions of the TUG. This study aimed to investigate the concurrent validity and test-retest reliability of the “up&go app,” which records and processes data from sensors embedded in a smartphone that is placed in the front pocket of the trousers. The results show excellent agreement between the “up&go” data and the comparator system, a previously validated medically certified lower-back inertial sensor system. Overall, the correlation of the 5 single repetitions was very strong. The results indicate that participants did not seem to walk at the same speed in each of the 5 test rounds, that is, they tended to walk slower during the first round. This trend probably shows a learning or accustoming effect. Similar observations have been made in previous studies [39,45] where participants started rather slowly and more carefully during the first attempt of motor

performance tests to follow the instructions precisely and to avoid errors [45]. This first round may therefore be comparable to a “trial run,” which is often performed in assessments before the actual measurement. Performing several repetitions of the test provides the opportunity to observe any changes in walking speed during the test and, if necessary, to interpret it clinically regarding physical capacity. For example, in more frail populations, a significant reduction in walking speed during the test could indicate fatigue. Furthermore, this approach might achieve a better approximation toward the “real” usual walking speed compared with measuring only 1 or 2 rounds. Using the app to monitor differences in the total test duration across several measurements would be a clinically relevant use case. But what would be considered a “relevant change”? Minimal important differences (MID) of >2 seconds for 1 TUG repetition were reported in studies observing populations with age-associated disorders and TUG baseline values around 20 seconds [46,47]. In another study observing older adults with hip osteoarthritis and a TUG baseline duration of 7 seconds, a sample that is comparable with this study, the MID is significantly lower with, 0.8 seconds [48]. This suggests that ceiling effects make it more difficult to screen for relevant changes in more robust individuals. The “up&go app,” however, should not only aim to screen for subjectively noticeable changes, but also for changes falling below the abovementioned MID thresholds. Looking at the Bland-Altman plots, a clinically relevant change could be assumed if the measured change is outside the limits of agreement. The limits identified in this study were about ± 3 seconds (LLoA, ULoA) for the total duration of 5 repetitions and about ± 1 second (LLoA, ULoA) for a single round. Especially in relatively fit, prefrail older adults,

a 5-repetition approach could be more sensitive to capture small changes and clinically more useful for early identification of declining physical capacity.

The test-retest reliability analysis showed excellent results. Looking at the average total test duration, we observed a shorter total test duration in the retest. On average, retests took place 6 days after the initial measurement, which makes an improved TUG test performance through training effects unlikely. A similar observation was described elsewhere [45]. A learning effect could account for the shorter duration of the retest. Since participants were already familiar with the test from the first measurement, they may have been more confident and faster during the retest. Other iTUG systems measuring at the height of the navel, (ICC=0.97) [49] and the lower back (ICC=0.9-0.96) [50] show slightly higher test-retest reliability values. This is to be expected, as the “up&go app” pocket position, specifically selected for the self-assessment purpose, is located more distant from the center of mass. Another reason could be that individuals wore different clothes with different types of pockets; these pockets may be tight or loose, which aggravates individual differences.

For future monitoring purposes, machine learning could be implemented to enable an in-app plausibility check. This might be a helpful feature to detect implausible changes in test duration (eg, >20% compared with previous test performances very recently).

Limitations

The sample was predominantly White, and participants were relatively fit, as demonstrated by the TUG performance. The average time needed to complete one repetition (9.7, SD 1.7 seconds; Table 1) is below the threshold of 12 seconds to distinguish normal versus below normal mobility [38]. The results are therefore limited in their transferability to more diverse cultural populations with greater mobility restrictions. For pragmatic reasons, we did not include a gold-standard reference measure, that is, optoelectronic measurement, limiting the results of this study accordingly. A segmental analysis of the “up&go app” data (sit-to-stand, walking, turning, and stand-to-sit) was not considered to be useful for self-monitoring. Furthermore, signal noise caused by the pocket-worn approach would be expected, as the smartphone moves back and forth in the trouser pocket in addition to the body movement during the test. Thus, the pocket-worn approach should be considered as a consumer-centered approach and not as a measurement system for segmental or kinematic analysis. For an accurate segmentation of the

test, the use of technology that is placed close to the center of mass would be recommended. The instruction to walk at “usual walking speed” leaves room for individual interpretation, hampering standardization. The extent to which the gait speed during the TUG test corresponds to an individual’s actual real-world gait speed was not assessed in this study. It is assumed that performing 5 repetitions and the familiar home setting enabled a greater approximation to the normal walking speed. However, an activity measurement of several days would be required for comparison [51].

Future Perspectives

The smartphone app used in this study could be suitable as a physical capacity screening and monitoring tool incorporating an instrumented approach for the widely used TUG test. It is designed to be used by older adults, but it could also be implemented by health care professionals to measure TUG total time in clinical settings.

As this study was aimed at validating the test algorithm, it is currently not possible to draw conclusions on the feasibility and user experience of older adults when operating the app, setting up and conducting the test independently. The app should therefore be examined and further developed in future cocreation studies with the target group.

The investigated app is designed to screen physical capacity, representing one of several health domains. In the future, it could be used as part of a comprehensive digital self-assessment, which should include other relevant risk factors such as physical activity, cognition, and vision in addition to physical capacity [24,52]. Self-guided early detection of risks in these domains could enable timely, more specific clinical diagnostics and the initiation of appropriate care interventions. Implementing the app as an upstream assessment within a digital training platform would allow adopting the dosage and supporting the selection of target-specific exercises to be tailored to the user’s individual level of physical capacity.

Conclusion

The results show excellent concurrent validity and test-retest reliability of the pocket-worn iTUG approach with 5 repetitions. The “up&go app” could be suitable as a self-screening and self-monitoring of physical capacity for older adults at home. It provides a smartphone-based approach to accurately measure the total duration of TUG. This novel approach offers the potential for older adults to take an active role in their health management and preclinical risk detection.

Acknowledgments

This publication is based on data from the study “Smart Aging in Community Contexts: Testing Intelligent Assistive Systems for Self-regulation and Co-regulation under Real-Life Conditions” (SMART-AGE). The study is funded by the Carl-Zeiss-Foundation. We thank all project members from the SMART-AGE project and the TSG Research Lab for their contribution. In particular, the authors extend thanks to Carlo Tacconi, Jan Spielmann, Jan Mayer, Ronny Bergquist, Sarah Zimmermann, Tobias Eckert, Claudia Hellmund, Franziska Kramer-Gmeiner, Meike Steinhilber, Anna-Lena Schubert and Alica Mertens.

Authors’ Contributions

MJB, CB, and KGO conceptualized the study. MJB, AB, EL, and MS took part in data collection. MJB and SM took part in the data analysis. MJB drafted the manuscript, JK, SM, CB, and KGO took part in a preliminary process of reviewing and editing the manuscript. All authors participated in the subsequent editing process, read the entire manuscript, and agreed to the published version.

Conflicts of Interest

SM holds a share of the mHealth Technologies srl company. None of the other authors declared conflicts of interest. The “up&go app” is freely available in the Google Play and Apple App Store, in German speaking countries.

References

1. Pew Research Center. Share of those 65 and older who are tech users has grown in the past decade. 2022. URL: <https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/> [Accessed 2024-10-09]
2. Chen C, Ding S, Wang J. Digital health for aging populations. *Nat Med*. Jul 2023;29(7):1623-1630. [doi: [10.1038/s41591-023-02391-8](https://doi.org/10.1038/s41591-023-02391-8)] [Medline: [37464029](https://pubmed.ncbi.nlm.nih.gov/37464029/)]
3. Fotteler M, Risch B, Gaugisch P, et al. Obstacles to using assistive technology for older adults - results from a focus group analysis. *Stud Health Technol Inform*. May 27, 2021;281:994-998. [doi: [10.3233/SHTI210327](https://doi.org/10.3233/SHTI210327)] [Medline: [34042822](https://pubmed.ncbi.nlm.nih.gov/34042822/)]
4. Wilson SA, Byrne P, Rodgers SE, Maden M. A systematic review of smartphone and tablet use by older adults with and without cognitive impairment. *Innov Aging*. 2022;6(2):igac002. [doi: [10.1093/geroni/igac002](https://doi.org/10.1093/geroni/igac002)] [Medline: [35243008](https://pubmed.ncbi.nlm.nih.gov/35243008/)]
5. Malani P, Kullgren J, Solway E, et al. National Poll on Healthy Aging: Mobile Health App Use Among Older Adults. 2022.
6. United Nations, Department of Economic and Social Affairs, Population Division. World population ageing 2019: highlights. [(ST/ESA/SER.A/430)]. 2019.
7. World Health Organization. Decade of healthy ageing: plan of action. 2020. URL: <https://www.who.int/publications/m/item/decade-of-healthy-ageing-plan-of-action> [Accessed 2024-10-09]
8. Amuthavalli Thiyagarajan J, Mikton C, Harwood RH, et al. The UN Decade of healthy ageing: strengthening measurement for monitoring health and wellbeing of older people. *Age Ageing*. Jul 1, 2022;51(7):afac147. [doi: [10.1093/ageing/afac147](https://doi.org/10.1093/ageing/afac147)] [Medline: [35776669](https://pubmed.ncbi.nlm.nih.gov/35776669/)]
9. Beard JR, Officer A, de Carvalho IA, et al. The World report on ageing and health: a policy framework for healthy ageing. *Lancet*. May 2016;387(10033):2145-2154. [doi: [10.1016/S0140-6736\(15\)00516-4](https://doi.org/10.1016/S0140-6736(15)00516-4)]
10. Honvo G, Sabico S, Veronese N, et al. Measures of attributes of locomotor capacity in older people: a systematic literature review following the COSMIN methodology. *Age Ageing*. Oct 28, 2023;52(Suppl 4):iv44-iv66. [doi: [10.1093/ageing/afad139](https://doi.org/10.1093/ageing/afad139)] [Medline: [37902521](https://pubmed.ncbi.nlm.nih.gov/37902521/)]
11. Podsiadlo D, Richardson S. The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc*. Feb 1991;39(2):142-148. [doi: [10.1111/j.1532-5415.1991.tb01616.x](https://doi.org/10.1111/j.1532-5415.1991.tb01616.x)] [Medline: [1991946](https://pubmed.ncbi.nlm.nih.gov/1991946/)]
12. Coelho-Junior HJ, Rodrigues B, Gonçalves I de O, Asano RY, Uchida MC, Marzetti E. The physical capabilities underlying timed “Up and Go” test are time-dependent in community-dwelling older women. *Exp Gerontol*. Apr 2018;104:138-146. [doi: [10.1016/j.exger.2018.01.025](https://doi.org/10.1016/j.exger.2018.01.025)] [Medline: [29410234](https://pubmed.ncbi.nlm.nih.gov/29410234/)]
13. Gatenio-Hefling O, Tzemah-Shahar R, Asraf K, Dilian O, Gil E, Agmon M. Revisiting the “Timed Up and Go” test: a 12-s cut-off can predict Hospitalization Associated Functional Decline in older adults. *Geroscience*. Jul 16, 2024. [doi: [10.1007/s11357-024-01280-3](https://doi.org/10.1007/s11357-024-01280-3)] [Medline: [39014130](https://pubmed.ncbi.nlm.nih.gov/39014130/)]
14. Ponciano V, Pires IM, Ribeiro FR, et al. Is the Timed-Up and Go Test feasible in mobile devices? A systematic review. *Electronics (Basel)*. 2020;9(3):528. [doi: [10.3390/electronics9030528](https://doi.org/10.3390/electronics9030528)]
15. Böttinger MJ, Labudek S, Schoene D, et al. “TiC-TUG”: technology in clinical practice using the instrumented timed up and go test-a scoping review. *Aging Clin Exp Res*. Apr 27, 2024;36(1):100. [doi: [10.1007/s40520-024-02733-7](https://doi.org/10.1007/s40520-024-02733-7)] [Medline: [38676844](https://pubmed.ncbi.nlm.nih.gov/38676844/)]
16. Milosevic M, Jovanov E, Milenkovic A. Quantifying timed-up-and-go test: A smartphone implementation. Presented at: 2013 IEEE International Conference on Body Sensor Networks (BSN 2013): Cambridge, MA, USA; May 6-9, 2013. [doi: [10.1109/BSN.2013.6575478](https://doi.org/10.1109/BSN.2013.6575478)]
17. Mellone S, Tacconi C, Chiari L. Validity of a Smartphone-based instrumented Timed Up and Go. *Gait Posture*. May 2012;36(1):163-165. [doi: [10.1016/j.gaitpost.2012.02.006](https://doi.org/10.1016/j.gaitpost.2012.02.006)] [Medline: [22421189](https://pubmed.ncbi.nlm.nih.gov/22421189/)]
18. Bergquist R, Nerz C, Taraldsen K, et al. Predicting advanced balance ability and mobility with an instrumented Timed Up and Go Test. *Sensors (Basel)*. Sep 3, 2020;20(17):4987. [doi: [10.3390/s20174987](https://doi.org/10.3390/s20174987)] [Medline: [32899143](https://pubmed.ncbi.nlm.nih.gov/32899143/)]

19. Bergquist R, Vereijken B, Mellone S, Corzani M, Helbostad JL, Taraldsen K. App-based self-administrable clinical tests of physical function: development and usability study. *JMIR Mhealth Uhealth*. Apr 27, 2020;8(4):e16507. [doi: [10.2196/16507](https://doi.org/10.2196/16507)] [Medline: [32338616](https://pubmed.ncbi.nlm.nih.gov/32338616/)]
20. Fudickar S, Pauls A, Lau S, et al. Measurement system for unsupervised standardized assessments of Timed Up and Go Test and 5 Times Chair Rise Test in community settings—a usability study. *Sensors (Basel)*. Jan 19, 2022;22(3):731. [doi: [10.3390/s22030731](https://doi.org/10.3390/s22030731)] [Medline: [35161478](https://pubmed.ncbi.nlm.nih.gov/35161478/)]
21. Ortega-Bastidas P, Gómez B, Aqueveque P, Luarte-Martínez S, Cano-de-la-Cuerda R. Instrumented Timed Up and Go Test (iTUG)—more than assessing time to predict falls: a systematic review. *Sensors (Basel)*. Mar 24, 2023;23(7):3426. [doi: [10.3390/s23073426](https://doi.org/10.3390/s23073426)] [Medline: [37050485](https://pubmed.ncbi.nlm.nih.gov/37050485/)]
22. Mellone S, Tacconi C, Schwickert L, Klenk J, Becker C, Chiari L. Smartphone-based solutions for fall detection and prevention: the FARSEEING approach. *Z Gerontol Geriat*. Dec 2012;45(8):722-727. [doi: [10.1007/s00391-012-0404-5](https://doi.org/10.1007/s00391-012-0404-5)]
23. Coni A, Ancum JMV, Bergquist R, et al. Comparison of standard clinical and instrumented physical performance tests in discriminating functional status of high-functioning people aged 61-70 years old. *Sensors (Basel)*. Jan 22, 2019;19(3):449. [doi: [10.3390/s19030449](https://doi.org/10.3390/s19030449)] [Medline: [30678268](https://pubmed.ncbi.nlm.nih.gov/30678268/)]
24. Böttinger MJ, Litz E, Gordt-Oesterwind K, et al. Co-creating a digital life-integrated self-assessment for older adults: user experience study. *JMIR Aging*. Sep 26, 2023;6:e46738. [doi: [10.2196/46738](https://doi.org/10.2196/46738)] [Medline: [37751274](https://pubmed.ncbi.nlm.nih.gov/37751274/)]
25. Soubra R, Chkeir A, Novella JL. A systematic review of thirty-one assessment tests to evaluate mobility in older adults. *Biomed Res Int*. 2019;2019:1354362. [doi: [10.1155/2019/1354362](https://doi.org/10.1155/2019/1354362)] [Medline: [31321227](https://pubmed.ncbi.nlm.nih.gov/31321227/)]
26. Goldsack JC, Coravos A, Bakker JP, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *NPJ Digit Med*. 2020;3:55. [doi: [10.1038/s41746-020-0260-4](https://doi.org/10.1038/s41746-020-0260-4)] [Medline: [32337371](https://pubmed.ncbi.nlm.nih.gov/32337371/)]
27. Kottner J, Audigé L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol*. Jan 2011;64(1):96-106. [doi: [10.1016/j.jclinepi.2010.03.002](https://doi.org/10.1016/j.jclinepi.2010.03.002)] [Medline: [21130355](https://pubmed.ncbi.nlm.nih.gov/21130355/)]
28. Federal Institute for Drugs and Medical Devices. German clinical trials register: SMART aging in community contexts: testing intelligent assistive systems for self-regulation and co-regulation under real-life conditions (SMART-AGE). 2024. URL: <https://drks.de/search/en/trial/DRKS00034316> [Accessed 2024-12-13]
29. Skevington SM, Lotfy M, O'Connell KA, WHOQOL Group. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Qual Life Res*. Mar 2004;13(2):299-310. [doi: [10.1023/B:QURE.0000018486.91360.00](https://doi.org/10.1023/B:QURE.0000018486.91360.00)] [Medline: [15085902](https://pubmed.ncbi.nlm.nih.gov/15085902/)]
30. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. *J Pers Assess*. Feb 1985;49(1):71-75. [doi: [10.1207/s15327752jpa4901_13](https://doi.org/10.1207/s15327752jpa4901_13)] [Medline: [16367493](https://pubmed.ncbi.nlm.nih.gov/16367493/)]
31. Groll DL, To T, Bombardier C, Wright JG. The development of a comorbidity index with physical function as the outcome. *J Clin Epidemiol*. Jun 2005;58(6):595-602. [doi: [10.1016/j.jclinepi.2004.10.018](https://doi.org/10.1016/j.jclinepi.2004.10.018)] [Medline: [15878473](https://pubmed.ncbi.nlm.nih.gov/15878473/)]
32. Reitan RM. Trail Making Test. Reitan Neuropsychology Laboratory; 1992.
33. Gale TM, Larner AJ. Six-item cognitive impairment test (6CIT). In: Larner AJ, editor. *Cognitive Screening Instruments*. Springer; 2017:241-253. [doi: [10.1007/978-3-319-44775-9_11](https://doi.org/10.1007/978-3-319-44775-9_11)] ISBN: 978-3-319-44774-2
34. Kempen GIJM, Yardley L, van Haastregt JCM, et al. The Short FES-I: a shortened version of the falls efficacy scale—international to assess fear of falling. *Age Ageing*. Jan 2008;37(1):45-50. [doi: [10.1093/ageing/afm157](https://doi.org/10.1093/ageing/afm157)] [Medline: [18032400](https://pubmed.ncbi.nlm.nih.gov/18032400/)]
35. 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*. Sep 2016;20(3):46-50. [doi: [10.20463/jenb.2016.09.20.3.7](https://doi.org/10.20463/jenb.2016.09.20.3.7)] [Medline: [27757387](https://pubmed.ncbi.nlm.nih.gov/27757387/)]
36. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as A measure of lower body strength in community-residing older adults. *Res Q Exerc Sport*. Jun 1999;70(2):113-119. [doi: [10.1080/02701367.1999.10608028](https://doi.org/10.1080/02701367.1999.10608028)] [Medline: [10380242](https://pubmed.ncbi.nlm.nih.gov/10380242/)]
37. Google Ireland Limited. Up&go (google play). 2024. URL: <https://play.google.com/store/apps/details?id=mht.tsg.upandgo&pli=1> [Accessed 2024-10-09]
38. Bischoff HA, Stähelin HB, Monsch AU, et al. Identifying a cut-off point for normal mobility: a comparison of the timed “up and go” test in community-dwelling and institutionalised elderly women. *Age Ageing*. May 2003;32(3):315-320. [doi: [10.1093/ageing/32.3.315](https://doi.org/10.1093/ageing/32.3.315)] [Medline: [12720619](https://pubmed.ncbi.nlm.nih.gov/12720619/)]
39. Kristensen MT, Ekdahl C, Kehlet H, Bandholm T. How many trials are needed to achieve performance stability of the Timed Up & Go test in patients with hip fracture? *Arch Phys Med Rehabil*. Jun 2010;91(6):885-889. [doi: [10.1016/j.apmr.2010.01.021](https://doi.org/10.1016/j.apmr.2010.01.021)] [Medline: [20510979](https://pubmed.ncbi.nlm.nih.gov/20510979/)]
40. Evans JD. *Straightforward Statistics for the Behavioral Sciences*. Brooks/Cole Pub. Co; 1996. ISBN: 0534231004

41. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. Feb 8, 1986;1(8476):307-310. [Medline: [2868172](#)]
42. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. Mar 1979;86(2):420-428. [doi: [10.1037//0033-2909.86.2.420](#)] [Medline: [18839484](#)]
43. Rosner B. *Fundamentals of Biostatistics*. 8th ed. Cengage Learning; 2016. ISBN: 130526892X
44. Tombaugh TN. Trail Making Test A and B: normative data stratified by age and education. *Arch Clin Neuropsychol*. Mar 2004;19(2):203-214. [doi: [10.1016/S0887-6177\(03\)00039-8](#)] [Medline: [15010086](#)]
45. Vartiainen MV, Rinne MB, Lehto TM, Pasanen ME, Sarajuuri JM, Alaranta HT. The test-retest reliability of motor performance measures after traumatic brain injury. *Adv Physiother*. Jan 2006;8(2):50-59. [doi: [10.1080/14038190600700195](#)]
46. Gallagher R, Marquez J, Osmotherly P. Clinimetric properties and minimal clinically important differences for a battery of gait, balance, and cognitive examinations for the Tap Test in idiopathic normal pressure hydrocephalus. *Neurosurgery*. Jun 1, 2019;84(6):E378-E384. [doi: [10.1093/neuros/nyy286](#)] [Medline: [30010977](#)]
47. Braun T, Thiel C, Schulz RJ, Grüneberg C. Responsiveness and interpretability of commonly used outcome assessments of mobility capacity in older hospital patients with cognitive spectrum disorders. *Health Qual Life Outcomes*. Mar 1, 2021;19(1):68. [doi: [10.1186/s12955-021-01690-3](#)] [Medline: [33648508](#)]
48. 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*. May 2011;41(5):319-327. [doi: [10.2519/jospt.2011.3515](#)] [Medline: [21335930](#)]
49. Yamada S, Aoyagi Y, Yamamoto K, Ishikawa M. Quantitative evaluation of gait disturbance on an instrumented Timed Up-and-go Test. *Aging Dis*. Feb 2019;10(1):23-36. [doi: [10.14336/AD.2018.0426](#)] [Medline: [30705765](#)]
50. Miller Koop M, Ozinga SJ, Rosenfeldt AB, Alberts JL. Quantifying turning behavior and gait in Parkinson's disease using mobile technology. *IBRO Rep*. Dec 2018;5:10-16. [doi: [10.1016/j.ibror.2018.06.002](#)] [Medline: [30135951](#)]
51. Warmerdam E, Hausdorff JM, Atrsaei A, et al. Long-term unsupervised mobility assessment in movement disorders. *Lancet Neurol*. May 2020;19(5):462-470. [doi: [10.1016/S1474-4422\(19\)30397-7](#)] [Medline: [32059811](#)]
52. 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*. Feb 1, 2014;14:14. [doi: [10.1186/1471-2318-14-14](#)] [Medline: [24484314](#)]

Abbreviations

ICC: intraclass correlation coefficient
iTUG: instrumented Timed Up and Go Test
LLoA: lower level of agreement
MID: minimal important difference
TUG: Timed Up and Go Test
ULoA: upper level of agreement
UN: United Nations
WHO: World Health Organization

Edited by Abdul-Fatawu Abdulai; peer-reviewed by Hikmet Kocaman, Silvia Rego; submitted 18.10.2024; final revised version received 13.12.2024; accepted 02.01.2025; published 21.03.2025

Please cite as:

Böttinger MJ, Mellone S, Klenk J, Jansen CP, Stefanakis M, Litz E, Bredenbrock A, Fischer JP, Bauer JM, Becker C, Gordt-Oesterwind K

A Smartphone-Based Timed Up and Go Test Self-Assessment for Older Adults: Validity and Reliability Study

JMIR Aging 2025;8:e67322

URL: <https://aging.jmir.org/2025/1/e67322>

doi: [10.2196/67322](#)

© Melissa Johanna Böttinger, Sabato Mellone, Jochen Klenk, Carl-Philipp Jansen, Marios Stefanakis, Elena Litz, Anastasia Bredenbrock, Jan-Philipp Fischer, Jürgen M Bauer, Clemens Becker, Katharina Gordt-Oesterwind. Originally published in *JMIR Aging* (<https://aging.jmir.org>), 21.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Aging*, is properly cited. The complete

bibliographic information, a link to the original publication on <https://aging.jmir.org>, as well as this copyright and license information must be included.