Properties of the ‘Timed Up and Go’ Test: More than Meets the Eye

Talia Herman a Nir Giladi a, c Jeffrey M. Hausdorff a, b, d

a Laboratory for Gait Analysis and Neurodynamics, Movement Disorders Unit, Department of Neurology, Tel-Aviv Sourasky Medical Center, Departments of b Physical Therapy, and c Neurology, Sackler Faculty of Medicine, Tel-Aviv University, Tel Aviv, Israel; d Harvard Medical School, Boston, Mass., USA

Key Words
Falls · Physical performance · Cognitive function · Balance · Gait · Aging · Mobility

Abstract
Background: The ‘timed up and go’ test (TUG) is a simple, quick and widely used clinical performance-based measure of lower extremity function, mobility and fall risk. We speculated that its properties may be different from other performance-based tests and assessed whether cognitive function may contribute to the differences among these tests in a cohort of healthy older adults. Objective: To evaluate psychometric properties of the TUG in healthy older adults in comparison to the Berg balance test (BBT) and the Dynamic Gait Index (DGI). Methods: The TUG, DGI and BBT were assessed in 265 healthy older adults (76.4 ± 4.3 years; 58.3% women) who participated in a 3-year prospective study. The Mini-Mental State Examination, digit span and verbal fluency measured cognitive function. The one-sample Kolmogorov-Smirnov test evaluated deviations from a normal distribution and Pearson’s correlation coefficients quantified associations. Results: The mean scores of the BBT, DGI and TUG were: 54.0 ± 2.4, 22.8 ± 1.5, 9.5 ± 1.7 s, respectively. The BBT and the DGI were not normally distributed (p < 0.001), but the TUG was (p = 0.713). The TUG times were mildly associated (p < 0.01) with digit span and verbal fluency and were related to future falls, while the BBT and the DGI were not. Conclusions: The TUG appears to be an appropriate tool for clinical assessment of functional mobility even in healthy older adults. It does not suffer from ceiling effect limitations, is normally distributed and is apparently related to executive function. The BBT and the DGI do not share these beneficial properties. Perhaps the transferring and turning components of the TUG help to convert this relatively simple motor task into a more complex measure that also depends on cognitive resources.

Introduction

The ‘timed up and go’ test (TUG) [1] is a simple, quick and widely used clinical performance-based measure of lower extremity function, mobility and fall risk. The TUG has been studied in elderly populations [1–8] and in various pathological conditions such as in patients with Parkinson’s disease [9, 58] (both ‘off’ and ‘on’ medication [10]), in patients with amyotrophic lateral sclerosis [11], in patients who are post-stroke [12], and in patients with orthopedic disturbances [13–15]. Numerous investigations have utilized the TUG as an outcome measure and demonstrated sensitivity to a variety of therapeutic interventions. Owing in part to its ease of use, association with fall risk and sensitivity, the American Geriatrics Society, the British Geriatrics Society [16], and the Society of Nordic Geriatricians [17], among others, recommend using The TUG as a screening test for fall risk.
The test procedure for the TUG is relatively simple. Subjects are asked to stand up from a standard chair (seat height between 44 and 47 cm), walk a distance of 3 m (marked on the floor) at a comfortable pace, turn, walk back and sit down. Subjects are permitted to use routine walking aids and are instructed not to use their arms to stand up. No physical assistance is given. The time to complete the test is measured with a stopwatch. Timing commences on the command ‘go’ and stops when the subject’s back is positioned against the back of the chair after sitting down. Usually the task is performed twice. Shorter times indicate better performance. Several studies have adopted a modified version of the test in which subjects are asked to walk as fast as they can, while ensuring safety [18, 19].

Some of the psychometric properties of the TUG have been reported previously. Inter-rater reliability is very high among hospital in-patients [1] (i.e. ICC = 0.99) and community-dwelling older adults (i.e. ICC = 0.98) [8]. The TUG also possesses high test-retest reliability [1, 20], although in a large study of older persons only moderate test-retest was found [21]. The TUG was able to correctly identify fallers and non-fallers (87% sensitivity and specificity), both as a single test or when subjects performed another test at the same time: subtraction (cognitive task) or carrying a full glass of water (manual task) [8]. It has been suggested that a cutoff point of 13.5 s can serve as a threshold for identifying persons with an increased risk of falling [8, 16]. Consistent with this, persons with vestibular disorders who took longer than 13.5 s to perform the TUG were 3.7 times more likely to have fallen in the previous 6 months [22, 23]. In contrast, a slightly lower cutoff point of 12 s has been applied to identify normal mobility in 413 community-dwelling elderly [24] and to differentiate fallers from non-fallers [25], while some have suggested that a 15-second threshold increases sensitivity (while providing insufficient specificity) [22]. In a cohort of 110 consecutive fall clinic patients, it was also indicated that 15 s in the TUG was the optimal cutoff point for identifying those with a high risk of falling [26]. In contrast, the TUG was statistically associated with a history of falls in men, but not in women, in a large cohort of 974 elderly with a history of falls. This study concluded that the ability to classify fallers with the TUG is poor [27]. The TUG has been also shown to be sensitive to interventions both in healthy community-living older adults [28, 29] and in patients with specific diseases, e.g. Parkinson’s disease [10, 30, 31].

While the TUG is one of the more commonly used clinical tools for quantifying gait, dynamic balance abilities and fall risk, several other functional performance-based tests have been developed to assess these parameters in various conditions. The Berg balance test (BBT) [3] and the Dynamic Gait Index (DGI) [32] are two tests that are also widely applied. All three of these clinical tests measure mobility and balance, but each has a unique focus. Hence, we speculated that their psychometric properties might be different. For example, a few previous studies reported ceiling effects in the BBT and the DGI in healthy older adults [19, 33, 34]; however, it is not clear if similar limitations exist for the TUG. The purpose of the present study was to investigate whether the TUG suffers from ceiling or floor effects in a relatively healthy cohort of older adults and to compare its properties to the BBT and the DGI. A second purpose was to examine the relationship between these three tests of mobility to cognitive abilities, mainly executive function, a factor also related to fall risk [35, 59]. Different associations between the three mobility tasks and executive function might explain test-specific properties.

**Methods**

**Subjects**

This report is based on the first year assessment and the first year follow-up of a cohort of 265 older healthy adults who were participating in a 3-year prospective study designed to assess the relationship between gait performance, postural control and cognitive function. Details and initial findings regarding dual tasking effects on gait and near falls have been reported elsewhere [36–38]. Briefly, the subjects were recruited from local senior centers, via flyers, lectures on this topic, advertising and word of mouth. An initial structured phone screening consisting of general health history was used. Eligible subjects were invited to participate if they were between 70 and 90 years of age, were community-dwelling older adults who walked independently (i.e. without a walking aid) and were free from disease likely to directly impact gait (e.g. vestibular, orthopedic, and neurological disease). Subjects were excluded if they had significant pain while walking, acute illness, brain surgery, major depression, history of stroke or if they scored less than 25 on the Mini-Mental State Examination (MMSE) [39]. The study was approved by the Human Studies Committee of the Tel-Aviv Sourasky Medical Center and informed written consent was obtained from all participants.

All subjects underwent a thorough evaluation including neurological examination, cognitive and clinical assessment conducted by a certified physical therapist. Medical history and status, demographic information and fall history during the 12-month period prior to enrolling in the study were obtained. Subjects completed forms that described their medical history and all prescription medications. These were reviewed to complete the Charlson comorbidity index, which quantified disease burden (higher scores indicate greater comorbidity) [40]. In addition, the motor portion (part III) of the unified Parkinson’s disease rating scale (UPDRS) quantified extra-pyramidal signs,
sometimes seen in older adults. Lower scores reflected fewer symptoms [41].

Assessment of Cognitive Function and Affect
In addition to the MMSE, a screening tool for dementia and general measure of cognitive function, subjects completed several tests of cognitive ability and executive function, including verbal fluency (VF) [42] and forward and backward digit span [43]. In the VF test, subjects were asked to recite out loud words that start with a predefined letter. The number of words generated in 1 min for three different Hebrew letters (bet/b; gimel/g; shin/sh) was timed using a stopwatch. These letters were chosen because they are commonly used in neuropsychological evaluation in Israel [42]. VF scores were achieved by calculating the sum of all words in the lists [42, 43]. In the digit span test, subjects were requested to repeat a series of digits which are read out loud by the examiner, in the same order (forward) or in reverse order (backwards) [43].

The Geriatric Depression Scale (GDS) [44] was used to assess the emotional wellbeing of the study participants. It includes 30 yes/no questions. A score of 0–10 is considered normal, 11–20 is a sign of mild depression and 21 and above suggests more severe depression. In addition, the state-trait anxiety inventory (STAI) was used to quantify the level of general and specific, current anxiety [45]. Higher scores indicate increased anxiety. The Activities-Specific Balance Confidence (ABC) Scale quantified fear of falling (scores closer to 100% indicate more confidence in daily activities and less fear of falling) [46, 47].

Performance-Based Measures of Balance and Mobility
As mentioned, all subjects performed the TUG without using an assistive device and the time was recorded using a stopwatch. Subjects performed the TUG twice, and the mean score was used for analysis [9]. The DGI was used to assess the subject’s ability to modify gait in response to changing task demands [32], e.g. walking while moving the head vertically or horizontally, walking while stepping over and around an obstacle, and stair climbing. A score lower than 19 points has been associated with impairment of gait and fall risk [32, 48]. The BBT was also administrated to evaluate balance and mobility [3]. Performance is rated on 14 different tasks, e.g. standing with eyes closed, tandem standing, single-leg stand, reaching, 360° turning and stepping. A score lower than 45 has been associated with minimal disease burden among the cohort. MMSE mean score was 28.7 ± 1.2 indicating that all subjects did not show signs of dementia or marked cognitive impairment. The GDS mean score was 5.3 ± 4.7 and within normal ranges. The mean UPDRS motor score (part III) was 4.1 ± 2.8; ranging from 0 to 15. The mean scores of STAI and the trait-anxiety inventory were 32.1 ± 10.1 and 33.8 ± 8.6, respectively, reflecting mild anxiety levels. Levels of fear of falling, as expected, were also within normal ranges and the mean ABC score was 92.0 ± 9.7%. The mean comfortable preferred walking speed was 1.3 ± 0.2 m/s.

Prospective Assessment of Falls
After the initial baseline assessment, data on falls were collected from all participants using a calendar that subjects filled out daily and returned every month by mail. For this study, a fall was defined as unintentionally coming to rest on the ground or other lower level not as a result of a major intrinsic event (e.g. myocardial infarction, stroke) or an external hazard (e.g. hit by a vehicle) [21]. Participants were instructed to complete a falls calendar by signing “X” for each fall on the day it occurred or mark “/” for everyday without a fall, and return it to the gait-lab via paid and pre-addressed envelopes. This approach has been well-validated for use in epidemiologic cohort studies [49, 50]. All subjects were instructed to keep the calendar in a convenient place (e.g. posted on a refrigerator) and to record the number of falls that occurred immediately after any event or at the end of each day [51]. Research staff closely monitored the return of the fall calendars. More than 80% of all subjects filled out and routinely returned these calendars; the rest were contacted via phone by a research assistant and missing data were collected. In general, subjects were contacted within a month in order to collect any missing data.

Statistical Analysis
Histograms and frequency distributions were constructed to evaluate normality and homogeneity of the distribution for the three tests of mobility. The one-sample Kolmogorov-Smirnov test evaluated deviations from a normal distribution. Pearson’s correlation coefficients were used to assess relationships between outcomes (for all outcomes, the results were similar if Spearman’s nonparametric correlations were assessed instead). All data were analyzed using the SPSS, version 15.0. A significance level of 0.05 was used.

Results

Subject Characteristics
The mean age of the 265 participants was 76.4 ± 4.3 years (range: 70–90) and 58.3% were women. As anticipated, among these relatively healthy elderly, the Charlson comorbidity index was low 0.82 ± 1.0, reflecting minimal disease burden among the cohort. MMSE mean score was 28.7 ± 1.2 indicating that all subjects did not show signs of dementia or marked cognitive impairment. The GDS mean score was 5.3 ± 4.7 and within normal ranges. The mean UPDRS motor score (part III) was 4.1 ± 2.8; ranging from 0 to 15. The mean scores of STAI and the trait-anxiety inventory were 32.1 ± 10.1 and 33.8 ± 8.6, respectively, reflecting mild anxiety levels. Levels of fear of falling, as expected, were also within normal ranges and the mean ABC score was 92.0 ± 9.7%. The mean comfortable preferred walking speed was 1.3 ± 0.2 m/s.

The TUG and Other Clinical Measures
The mean TUG score was 9.5 ± 1.7 s, ranging from 5.4 to 15.6. The mean BBT score was 54.1 ± 2.4, ranging from 45 to 56. The mean DGI score was 22.8 ± 1.5, ranging from 17 to 24. All these scores were near the maximal score and indicative of good balance and mobility. As seen in figure 1, the BBT and the DGI were not normally distributed (p < 0.001), but the TUG was (p = 0.713). Although the TUG, the DGI and the BBT all measure balance and mobility, they were only moderately correlated with one another (table 1) or with other measures assessed (table 2).
The TUG and Cognitive Function

The TUG times were mildly negatively correlated with MMSE \((r = –0.19; p = 0.002)\), digit span \((r = –0.18; p = 0.005)\) and verbal fluency \((r = –0.21; p = 0.001)\), while the BBT and the DGI were not correlated with these cognitive function measures \((p > 0.18)\). Mean forward digit span score was 9.16 ± 2.4, ranging from 3 to 15 digits. Mean backward digit span score was 6.04 ± 2.4, ranging from 1 to 13 digits. The mean VF score was 33.5 ± 12.4, ranging from 5 to 69 words. We further explored this association by comparing subjects with relatively poor versus relatively good VF performance. Sixty-nine subjects (26%) had VF scores in the upper quartile, with the mean number of words generated being 49.5 ± 6.6. Sixty-two subjects (23%) were in the lower quartile VF scores, with the mean number of words generated being 18.1 ± 4.9.

When comparing the TUG times in relation to VF performance, significant differences were observed \((p = 0.011)\), between the upper (better) quartile and lower (worse) quartile of VF. In contrast, this was not found in the DGI or BBT scores (fig. 2). Similar associations between the TUG and cognitive function were found if subjects were ranked based on their performance on the digit span (backward or total, \(p < 0.01)\). In contrast, the TUG was not associated with forward digit span \((p = 0.21)\) and the DGI and the BBT were not associated with digit span (forward, backward or total) when subjects in the lower and upper performance quartiles were compared.

The TUG Times and Falls

During the 12-month follow-up period, subjects who fell twice or more (multiple fallers; \(n = 26)\) took longer \((p = 0.035)\) to complete the TUG at baseline (10.3 ± 1.9 s),
compared to non-fallers (n = 201; 9.5 ± 1.7 s), while the BBT and the DGI were not significantly related to falls (p > 0.05). At baseline, only 5 subjects took more than 13.5 s to complete the TUG. The association between falls (single and multiple) during the 12-month follow-up period and the TUG times above 13.5 s at baseline was not significant (p > 0.544).

### Discussion

A priori, one could argue that the DGI and the BBT are relatively complex and comprehensive tests and, hence, less prone to floor or ceiling effects compared to the TUG test. The DGI challenges the subject in multiple ways (e.g. stair climbing, walking while rotating the head). Similarly, the BBT not only assesses quiet standing with eyes open (a relatively simple task), it also measures 13 other aspects of balance and mobility including single-legged stance, functional reach and tandem standing abilities. In contrast, the TUG (at least at first glance) is a relatively simple test. The results of the present study indicate that these assumptions, at least in healthy older adults, are not completely correct. The TUG was normally distributed and did not suffer from ceiling or floor effects in healthy older adults. Conversely, the two other commonly used tests of mobility and fall risk, the BBT and the DGI, suffered from ceiling effects, consistent with previous findings in highly functioning older adults [19, 33, 34]. We also observed small, but significant, associations between the TUG and tests of executive function that were not seen when exam-
In the present study, the authors and those with early functional limitations and disability successfully detected (p = 0.002) differences in performance between the patients with relatively high function and those with early functional limitations and disability. Consistent with the present findings, this led the authors of that study to conclude that the TUG is sensitive to early changes in functional status.

Despite the apparent advantages of the TUG over the BBT and the DGI, the present findings also raise some questions which require future study. For example, closer comparison of these three tests suggests that while the focus is different, all three are compromised of multiple subtasks that demand a variety of different motor and cognitive resources. Even the relatively ‘static’ BBT includes the assessment of turning, for example. Differences in characteristics among the tests may be related not only to the fine structure of the various components, but also to the fact that the TUG outcome measure is a continuous one (time, in seconds), while the BBT and the DGI scores are the sum of ordinal scales. In addition, the present results underscore the notion that different TUG cutoff thresholds may be appropriate under different circumstances. In the healthy elderly, a TUG time above 13.5 s may be a good indicator of poor mobility, but might not be the ideal value for providing an early marker of fall risk.

To summarize, our findings suggest that the TUG is moderately related to cognitive ability, in particular executive function, while the BBT and the DGI are not. Although the clinical relevance of this relatively mild association may be limited in healthy elderly, it may take on a more significant meaning in populations that have impaired cognitive function and during rehabilitation (e.g. post-stroke). The transferring and turning components of the TUG may help convert this relatively simple motor task into a more complex measure of mobility that also depends on cognitive resources. Still, the present findings challenge researchers to better understand why executive function is associated with the TUG, but not the BBT or the DGI. In addition, these results suggest that the TUG may have certain advantages over other widely used tests that may be especially relevant when studying or detecting early decline in otherwise healthy older adults.

Acknowledgements

This work was supported in part by the National Institute on Aging (AG14100) and by the European Commission in the context of FP6 projects DAPHNet, fet-018474-2, and SENSATION-AAA, info-ist-045622. The authors are indebted to the participants and staff of the ‘Holchim Rachok’ project for their invaluable contribution to this project, especially Marina Brozgol, Noit Inbar-Borovsky, Aner Weiss and Leor Gruendlinger. Portions of this work were presented at the annual meeting of the American Geriatrics Society, May 2008.


