Prognostic validity of the Timed Up-and-Go test, a modified Get-Up-and-Go test, staff’s global judgement and fall history in evaluating fall risk in residential care facilities

ELLINOR NORDIN1, NINA LINDELÖF2,3, ERIK ROSENDAHLE, JANE JENSEN1, LILLEMOR LUNDIN-OILSSON1

1Department of Community Medicine and Rehabilitation, Physiotherapy, Umeå University, SE-901 87 Umeå, Sweden
2Department of Health Sciences, Physiotherapy Unit, Luleå University of Technology, SE-971 87 Luleå, Sweden
3Department of Community Medicine and Rehabilitation, Geriatric Medicine, Umeå University, SE-901 87 Umeå, Sweden

Address correspondence to: E. Nordin. Tel: +46 90 786 9429; Fax: +46 90 786 9267. Email: ellinor.nordin@physiother.umu.se

Abstract

Objectives: to evaluate and compare the prognostic validity relative to falls of the Timed Up-and-Go test (TUG), a modified Get-Up-and-Go test (GUG-m), staff’s judgement of global rating of fall risk (GLORF) and fall history among frail older people.

Design: cohort study, 6-month prospective follow-up for falls.

Participants: 183 frail persons living in residential care facilities in Sweden, mean age 84 years, 73% women.

Methods: the occurrence of falls during the follow-up period were compared to the following assessments at baseline: the TUG at normal speed; the GUG-m, a rating of fall risk scored from 1 (no risk) to 5 (very high risk); the GLORF, staff’s rating of fall risk as ‘high’ or ‘low’; a history of falls in the previous 6 months. These assessment tools were evaluated using sensitivity, specificity and positive and negative likelihood ratios (LR+ to rule in and LR− to rule out a high fall risk).

Results: 53% of the participants fell at least once. Various cut-off values of the TUG (12, 15, 20, 25, 30, 35, 40 s) and the GUG-m showed LR+ between 0.9 and 2.6 and LR− between 0.1 and 1.0. The GLORF showed an LR+ of 2.8 and an LR− of 0.6 and fall history showed an LR+ of 2.4 and an LR− of 0.6.

Conclusions: in this population of frail older people, staff judgement of their residents’ fall risk as well as previous falls both appear superior to the performance-based measures TUG and GUG-m in ruling in a high fall risk. A TUG score of less than 15 s gives guidance in ruling out a high fall risk but insufficient information in ruling in such a risk. The grading of fall risk by GUG-m appears of very limited value.

Keywords: fall risk, prognostic validity, likelihood, sensitivity, specificity, elderly

Introduction

Falls are the most common cause of injury in older age [1]. Older people living in residential care facilities experience falls three times as frequently as older people living in the community [2, 3]. Although falls can be caused by multiple factors, mobility impairments such as gait and balance disorders are among the most common predisposing factors [4, 5].

We need prognostic tools that help identify individuals with an increased risk of falling in order to take preventive action. A frequently used tool is the Timed Up-and-Go Test (TUG), recommended by the American Geriatrics Society, the British Geriatric Society and Nordic Geriatricians for screening for risk of falling [6, 7]. The TUG is a timed performance of getting up from a chair, walking 3 m, turning around, and walking back to sit down again [8]. The origin of this test was the Get-Up-and-Go Test (GUG), an observational rating of fall risk using a score from 1 to 5 [9]. The GUG was later criticised for lacking precision, as the intermediate scores were not clearly defined [8], and the TUG has been increasingly used over the years. However, our clinical experience indicates that a timed performance may conceal an unbalanced gait, and an individual may well transfer rapidly despite walking difficulties. Our observations are supported by recent findings that older patients with Alzheimer’s disease walk too fast, considering their frailty [10]. So, even though...
the original GUG was perceived as imprecise, a modification with descriptions of each grade could give more information about the risk of falling than the TUG.

There are aspects of fall risk that neither the TUG nor the GUG can measure, such as acute illness and the individual's capacity to interpret and cope with situations in everyday life. These factors have been shown to precipitate falls in residential care settings [2]. The staff in a residential facility possesses 24-h knowledge of their residents' potential fall risk, which involves both pre-disposing and precipitating factors. Therefore, their global assessment of fall risk could have the highest predictive validity in relation to falls. The staff's judgement of 'high' or 'low' risk has been found to predict falls [11], but this finding has not been externally validated. Previous falls, often considered a gold standard in predicting falls [12], may influence the staff's judgement and should therefore be validated alongside the global assessment.

A prognostic tool should distinguish those who are likely to fall from those who are not. A subjective rating of fall risk as 'high' or 'low', as reflected by the staff's judgement, is dichotomous by nature; however, for measures based on continuous or ordinal scales such as the TUG and the GUG, cut-off values must be chosen. Cut-off TUG values between 10 and 25 s have been reported to discriminate between individuals who have or have not fallen [13–17]. However, all but one of these studies were case-controlled and based on a retrospective registration of falls, and the only longitudinal study had a poor follow-up on falls [16]. These factors strongly limit the inferences that clinicians can draw using the TUG as a prognostic tool. Regarding the GUG, scores of 2 or 3 indicate a history of falls [18], or unsteadiness in balance [19]. Thus, the prognostic validity of the TUG and the GUG has not been sufficiently evaluated.

Our purpose was to evaluate and compare the prognostic validity of the TUG, a modified Get-Up-and-Go test (GUG-m), the staff's judgement as well as fall history in classifying fall risk among older people living in residential care facilities. Our hypothesis was that a subjective rating of fall risk is better than a timed assessment of mobility, and that the staff's judgement has the best predictive validity, equivalent to that of previous falls.

Methods

Settings and participants

Data were collected from residential care facilities in Umeå, Sweden. The participants derived from a group of 254 residents, who were 65 years of age or older, had Mini-Mental State Examination (MMSE) scores of 10 or more and had physician's approval to participate in the study. Of these 254 residents, 183 participants were able to perform a baseline TUG.

The study was approved by the Ethics Committee of the Medical Faculty of Umeå University. The participants or their proxies, when appropriate because of cognitive impairment, gave their consent of participation. The MMSE score of 10 or more was an inclusion criteria for participants to be able to communicate and interact in a meaningful way [20].

Assessments

In the TUG [8] and GUG-m assessments, an armchair of standard height was used and a distance of 3 m was marked with a line of tape on the floor. The starting position was sitting with hands resting on the arms of the chair. The participants crossed the line before turning around and walking back to sit down in the chair again. They were instructed to perform the TUG at their normal speed and they performed one trial before they were timed. The timing of the TUG started when the participant's back came off the back of the chair, and stopped when their buttocks touched the seat of the chair again.

The scoring of the GUG test [9] was modified (GUG-m) by describing each of the five steps of the scale as follows (for further description see appendix 1 in the supplementary data on the journal's website http://www.ageing.oxfordjournals.org):

1. No Fall Risk. Well-coordinated movements, without walking aid.
2. Low Fall Risk. Controlled but adjusted movements.
4. High Fall Risk. Supervision necessary.
5. Very High Fall Risk. Physical support or stand by physical support necessary.

Two physical therapists' simultaneous ratings of GUG-m scores (n = 2 × 20) resulted in 10 pairs of equal scores, nine pairs with a difference of 1 scale step and one pair with a difference of 2 scale steps [weighted kappa coefficient 0.55 (95% CI 0.35–0.76)].

The global rating of fall risk (GLORF) was scored by a licensed practical nurse or nurse’s aid with personal knowledge of the resident. She/he answered the question, “how do you judge the risk that Mr or Mrs X will fall within 6 months—high or low?” [11]. Two raters were asked independently (n = 2 × 138), and a total agreement was obtained in 82% of the ratings [kappa value 0.60 (95% CI 0.47–0.74)]. After scoring the GLORF the staff answered questions about the resident’s dependence in activities of daily life and fall history (‘Has Mr or Mrs X fallen in the previous 6 months?’).

Outcome

A fall was defined when a person unintentionally came to rest on the floor or ground, regardless of the cause and the consequences of the fall. Falls were registered for 6 months following the baseline assessments. On the basis of the results from a previous study, 6 months was considered a relevant time period in this population, in which functional mobility may deteriorate rapidly [21]. Falls were documented by the staff, on specific report forms that were part of an existing routine. To further improve the registration of falls, the participants’ regular medical charts were also reviewed.
E. Nordin et al.

Statistical analysis

Statistical analyses were performed using SPSS software, version 14.0 (SPSS Inc., Chicago, Illinois).

Finding a single cut-off value for fall risk based on analysis of a Receiver Operating Characteristic (ROC) curve, for example, is a sample-dependent process and as such is difficult to extrapolate beyond a particular study. Other ways to choose cut-points have been suggested [22]. We chose to evaluate several cut-off values, some of which are related to different levels of functional mobility. According to the TUG scores, 12 s is suggested as the upper limit of normal mobility in older women [23], and less than 20 s is reported to reflect independence in extended functional mobility skills [8, 23]. We arbitrarily chose to evaluate also 15, 25, 30, 35 and 40 s as cut-off levels. The predictive accuracy of the GUG-m was evaluated using the full scale and looking at each score as a potential cut-off level. ROC curves were evaluated to compare the area under the curves with the different assessment tools.

Prognostic accuracy, the ability of each measure to correctly classify each participant as having either a high or a low risk of falling during the follow-up period, was calculated by means of sensitivity (the proportion of subjects correctly classified with a high fall risk among those falling at least once) and specificity (the proportion of subjects correctly classified with a low fall risk among those not falling), as well as by combining those into Likelihood Ratios (LR) [24–26]. In relation to falls, the LRs tell us how much to increase or decrease the probability of a future fall. A positive LR (LR+) is calculated as the sensitivity divided by (1−specificity), whereas a negative LR (LR−) is calculated as (1−sensitivity) divided by the specificity [24–26]. The result will guide us in ruling in (LR+) or ruling out (LR−) a high fall risk. On the basis of a pre-test probability (in our context, the proportion of residents with at least one fall) a post-test probability can be calculated for an individual patient by the use of the LR [25, 26].

Results

The participants in the study had a mean age of 84.3 years and 73% of them were women. Their median TUG score was 25.5 s and median GUG-m score was 3 points. Of the 183 participants 71 residents (39%) had a high fall risk according to GLORF, of whom 54 (76%) had fallen in the previous 6 months. A total of 74 participants (40%) had a history of falls (Table 1).

Ninety-seven participants (53%) fell at least once during the follow-up period. The number of falls ranged from 0 to 26 per person. The fall rate was 4.0 per person-year. Both those who fell and those who did not fall had a range of scores for the TUG, GUG-m, GLORF and for fall history (Figure 1).

Prognostic Value

ROC areas under the curves ranged from 0.62 to 0.69 (Table 2). TUG cut-off scores of 12, 15 and 20 s as well as the GUG-m cut-off score of 1 resulted in a negative LR (LR−) of 0.5 or lower and had a high sensitivity, but a low specificity. TUG cut-off scores of 35 and 40 s resulted in a positive LR (LR+) of 2.6 and 2.2, respectively, but the sensitivity was low. None of the GUG-m cut-off scores had LR− of ≥ 2 (Table 2).

The GLORF had an LR+ of 2.8 but an LR− of 0.6. A fall history showed similar results. The GLORF and fall history showed a sensitivity of 56 and 58%, respectively, while the specificity was higher (Table 2).

With a pre-test probability of falls of 53% and an LR− of 0.1, the post-test probability of ruling out a high fall risk using a TUG value of less than 15 s was 90%. For a TUG value of 12 s or less, the post-test probability was 82%, and for a TUG value of less than 20 s, it was 64%. The post-test probability of ruling in a high fall risk was 75% for a TUG value of 35 s or more and 71% for a TUG value of 40 s or more. A GUG-m score of 1 had an LR− of 0.4, which gave a post-test probability of 69% in ruling out a high fall risk. The GLORF had an LR+ of 2.8, which gave a post-test probability of...
Evaluating fall risk in residential care facilities

probability of ruling in a high fall risk of 76%. With an LR$^+$ of 2.4 for fall history, this probability was 73%.

Discussion
The sensitivity and specificity of assessment tools are seldom reported in relation to fall registration based on a prospective design. In a recently published review [27] on fall risk assessment tools, only three prospective studies [11, 21, 28] were found in which predictive values were calculated for older people in long-term settings. We calculated LRS from data presented in these studies which evaluated two multifactorial assessment tools: Downton index [21] and Mobility Interaction Fall Chart [11, 28]. Despite initially promising predictive values (LR$^+$4.7 and LR$^-$0.2) [28] none of the

Figure 1. Distribution of the scores in relation to those who did not fall, to those who fell once, and to those who had multiple falls during the 6 months follow-up for falls: (A) The Timed Up-and-Go (TUG) test, bars represent the TUG scores arbitrarily grouped in seconds ($n = 173$). (B) The modified Get-Up-and-Go (GUG-m) scores ($n = 183$). (C) The staff global rating of fall risk (GLORF) ($n = 183$). (D) History of falls in previous 6 months ($n = 183$).
tools had a diagnostic value in ruling in a high fall risk in the
external validations [11, 21] but Downton index was useful
ruling out a high fall risk (LR<0.2) [21]. Additionally, staff
judgement and fall history were equally useful in ruling in a
high fall risk (LR>1) [11]. Our external validation provides
further evidence in favour of the prognostic value of staff
judgement in ruling in a high fall risk. Furthermore, the LR+
that we found for fall history is in accordance with reports
based on falls in the previous year for an older population
dwelling in the community [12]. We found individuals with no falls, single falls and
recurrent falls across the whole range of scores for each of the
evaluation tools. This illustrates the non-linear relationship
between mobility and fall risk in which behavioural and
environmental factors, among others, modify the occurrence
of falls at different levels of mobility [29]. One source of
the ambiguous results of the TUG scores in relation to falls is the 35% day-to-day variability of TUG scores in frail older people [30]. Although TUG scores of 35 s or more were found to rule in a high fall risk, TUG scores between 20 and 30 s neither ruled in nor ruled out such a risk. This time span has previously been reported as a ‘grey zone’ of functional skills, whereas TUG scores of 30 s or more reflect dependency on others for safe transfers [8]. Our results suggest that TUG has insufficient value in ruling in a high fall risk but give guidance to rule out such a risk.

The observational rating GUG-m was inferior to the timed
assessment of mobility and showed relatively poor inter-rater
reliability. So, despite the small additional guidance for some residents rated as having ‘no fall risk’, the prognostic validity of the GUG-m in fall risk assessment is very limited in a population of frail older people with assisted living.

When it comes to ruling in a high fall risk, the GLORF may take into account important precipitating factors, thereby encompassing a wider concept of physical mobility compared to the TUG or GUG-m. One fourth of those judged to have a high fall risk by the GLORF had no fall history in the previous 6 months, so the judgement was not simply based on the TUG or GUG-m. One fourth of those judged to have a high fall risk by the GLORF had no fall history in the previous 6 months, so the judgement was not simply based on the TUG or GUG-m. One fourth of those judged to have a high fall risk by the GLORF had no fall history in the previous 6 months, so the judgement was not simply based on the TUG or GUG-m.

We acknowledge certain limitations to the present study.
The data was collected in the residential facilities in a
Swedish city; thus it can be questioned whether results can be generalised to assisted living situations in other
countries. To enable practitioners to assess whether our
results may be applicable to other populations, we made an
effort to present extensive background characteristics of the
participants in this study. Nonetheless, whether our results

### Table 2. Predictive values of the Timed Up-and-Go Test (TUG), the Modified Get-Up-and-Go Test (GUG-m), the Global Rating of Fall Risk (GLORF) and a history of falls during a 6-month prospective follow-up for falls comparing one or more falls versus no falls

<table>
<thead>
<tr>
<th>Type of Measure</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>LR+, (95% CI)</th>
<th>LR−, (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TUG</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROC area0.60 0.61—0.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUG cut-off 12 s 96 (92–100)</td>
<td>13 (7–22)</td>
<td>1.1 (1.0–1.2)</td>
<td>0.2 (0.0–0.7)</td>
<td></td>
</tr>
<tr>
<td>TUG cut-off 15 s 96 (92–100)</td>
<td>32 (21–42)</td>
<td>1.4 (1.2–1.6)</td>
<td>0.1 (0.0–0.4)</td>
<td></td>
</tr>
<tr>
<td>TUG cut-off 20 s 79 (69–86)</td>
<td>32 (36–58)</td>
<td>1.5 (1.2–1.9)</td>
<td>0.5 (0.3–0.7)</td>
<td></td>
</tr>
<tr>
<td>TUG cut-off 25 s 62 (52–72)</td>
<td>62 (51–73)</td>
<td>1.6 (1.2–2.2)</td>
<td>0.6 (0.5–0.8)</td>
<td></td>
</tr>
<tr>
<td>TUG cut-off 30 s 49 (39–59)</td>
<td>72 (62–82)</td>
<td>1.8 (1.2–2.6)</td>
<td>0.7 (0.6–0.9)</td>
<td></td>
</tr>
<tr>
<td>TUG cut-off 35 s 36 (26–46)</td>
<td>86 (78–94)</td>
<td>2.6 (1.4–4.8)</td>
<td>0.7 (0.6–0.9)</td>
<td></td>
</tr>
<tr>
<td>TUG cut-off 40 s 26 (17–34)</td>
<td>89 (82–96)</td>
<td>2.2 (1.1–4.5)</td>
<td>0.8 (0.7–1.4)</td>
<td></td>
</tr>
<tr>
<td><strong>GUG-m</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROC area0.62 0.54–0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GUG-m cut-off 1, ‘No Risk’ 94 (87–97)</td>
<td>16 (10–26)</td>
<td>1.0 (1.0–1.2)</td>
<td>0.4 (0.2–0.9)</td>
<td></td>
</tr>
<tr>
<td>GUG-m cut-off 2, ‘Low Risk’ 62 (52–71)</td>
<td>60 (50–70)</td>
<td>1.6 (1.2–2.1)</td>
<td>0.6 (0.5–0.9)</td>
<td></td>
</tr>
<tr>
<td>GUG-m cut-off 3, ‘Some Risk’ 20 (19–37)</td>
<td>83 (75–91)</td>
<td>1.6 (0.9–2.8)</td>
<td>0.9 (0.7–1.0)</td>
<td></td>
</tr>
<tr>
<td>GUG-m cut-off 4, ‘High Risk’ 7 (2–12)</td>
<td>92 (86–98)</td>
<td>0.9 (0.3–2.4)</td>
<td>1.0 (0.9–1.1)</td>
<td></td>
</tr>
<tr>
<td><strong>GLORF</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROC area0.68 0.60–0.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘High Risk’ — ‘Low Risk’ 56 (46–65)</td>
<td>80 (71–87)</td>
<td>2.8 (1.8–4.5)</td>
<td>0.6 (0.4–0.7)</td>
<td></td>
</tr>
<tr>
<td><strong>History of Falls</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROC area0.66 0.58–0.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In previous 6 months: yes–no 58 (48–68)</td>
<td>76 (67–85)</td>
<td>2.4 (1.6–3.7)</td>
<td>0.6 (0.4–0.7)</td>
<td></td>
</tr>
</tbody>
</table>

* Confidence interval.

+ Positive likelihood ratio, interpretation of change in disease likelihood: >10, large; 5–10, moderate; 2–5, small; 1–2, no or tiny.

− Negative likelihood ratio, interpretation of change in disease likelihood: <0.1, large; 0.1–0.2, moderate; 0.2–0.5, small, >0.5–1, no or tiny. (Diagnostic test. User’s Guides to Medical Literature. Evidence-based medicine working group (on line). Available at: http://ebem.org/usersguides/diagnosis.html accessed October 26, 2007.)

— Receiver operating characteristic area under the curve (95% CI).
are representative of other populations needs to be validated in independent samples of older individuals.

In this population of frail older people, staff judgement of their residents’ fall risk as well as previous falls both appear superior to the performance-based measures TUG and GUG-m in ruling in a high fall risk. A TUG score of less than 15 s gives guidance in ruling out a high fall risk but provides insufficient information in ruling in such a risk. The grading of fall risk by GUG-m appears to be of very limited value.

Key points
- Fall risk assessments may guide clinicians in two directions, either in ruling in a high fall risk or in ruling out such a risk. However, a single cut-off score does not necessarily give guidance in both directions.
- There is a non-linear relationship between mobility and fall risk that makes TUG inconclusive in ruling in a high fall risk. A TUG value of less than 15 s gives guidance in ruling out such a risk.
- Staff knowledge is superior to performance-based measures of mobility and fall prevention should therefore be intensified when staff members, who know the resident well, judge the fall risk to be high, or when a person has had a fall.

Funding sources
This investigation was supported by grants from the Swedish Research Council (K2004-27KX-15041-01A and K2005-27X-15357-01A), the Swedish Council for Working Life and Social Research, the Aldrecentrum Vasterbotten, the Erik and Anne-Marie Detlof’s Foundation, Umeå University, the SJCKMS and Gun and Bertil Stohne’s Foundation.

Conflict of interest
There is no conflict of interest related to this work.

Informed consent
The study was approved by the Ethics Committee of the Medical Faculty, Umeå University: § 391/01 and 439/03. The study was approved by the Ethics Committee of the Medical Faculty, Umeå University: § 391/01 and 439/03. There is no conflict of interest related to this work.

Supplementary data
Supplementary data for this article are available online at http://ageing.oxfordjournals.org.

References

Received 27 April 2007; accepted in revised form 5 February 2008