The comparative ability of eight functional mobility tests for predicting falls in community-dwelling older people

ANNE TIEDEMANN1, HIROYUKI SHIMADA2, CATHERINE SHERRINGTON1, SUSAN MURRAY1, STEPHEN LORD2

1Prince of Wales Medical Research Institute, Barker St, Randwick NSW 2031, Australia
2Tokyo Metropolitan Institute of Gerontology, 35-2 Sakae-cho, Itabashi-ku, Tokyo 173-0015, Japan

Address correspondence to: A. Tiedemann. Tel: (+61) 2 9399 1060; Fax: (+61) 2 9399 1005. Email: a.tiedemann@powmri.edu.au

Abstract

Background: numerous tests have been suggested as fall risk indicators. However, the validity of these assessments has not been demonstrated in large representative samples of community-dwelling older people.

Objective: the objective of this study was to examine the comparative ability and clinical utility of eight mobility tests for predicting multiple falls in older community-dwelling people.

Methods: design—prospective cohort study; subjects—362 subjects aged 74–98 years; measurements—the sit-to-stand test with one and five repetitions, the pick-up-weight test, the half-turn test, the alternate-step test (AST), the six-metre-walk test (SMWT) and stair ascent and descent tasks. Falls were monitored for 1 year with fall calendars.

Results: in the 12-month follow-up period, 80 subjects (22.1%) suffered two or more falls. Multiple fallers performed significantly worse than non-multiple fallers in the sit-to-stand test with five repetitions (STS-5), the AST, the half-turn test, the SMWT and the stair-descent test. When dichotomised using cut-off points from receiver-operated characteristics (ROC) curve analyses, these tests demonstrated reasonable sensitivity and specificity in identifying multiple fallers. A principal components analysis identified only one factor underlying the mobility tests. Poor performances in two mobility tests, however, increased the risk of multiple falls more than poor performance in one test alone (ORs = 3.66, 95% CI = 1.44, 9.27 and 1.61, 95% CI = 0.62, 4.16 respectively).

Conclusions: the mobility tests appear to be measuring a similar ‘mobility’ construct. Based on feasibility and predictive validity, the AST, STS-5 and SMWTs were the best tests.

Keywords: accidental falls, mobility tests, mobility, balance, aged, elderly

Introduction

Falls are a serious public health issue, which contribute to substantial morbidity and mortality in older people [1]. Identification of risk factors and consequent intervention is an effective method of preventing falls in the elderly [2]. Previously, several tests have been developed for assessing mobility in older people, some of which have also been suggested as predictors of falls [3–10]. However, most tests have not been validated prospectively in large representative samples of community-dwelling older people, making their predictive accuracy uncertain.

The aims of this study were to determine the test/re-test reliabilities and the comparative abilities of eight mobility tests for predicting multiple falls in a large sample of older community-dwelling people. Multiple fallers were compared to non-multiple fallers as it has been reported that recurrent falls are more likely to indicate physiological impairments and chronic conditions [3, 11] and are therefore more clinically important. The tests were chosen because they had a short administration time and did not require specialised equipment, making them suitable for use in clinical settings.

Methods

Subjects

Participants were randomly selected from the membership database of a health insurance company as part of a falls prevention randomised controlled trial conducted between 1999 and 2002 [12]. Exclusion criteria included minimal English, blindness, Parkinson’s disease or a Short Portable Mental Status Questionnaire (SPMSQ) score <7 [13]. The
mobility tests were carried out at an acute care hospital, and transport was provided for people with mobility limitations. Four hundred subjects (137 men) aged 74–98 years (mean 80.4, SD = 4.5) from the control group met the inclusion criteria and completed the mobility tests. Based on a previous study [11], a sample size of 360 subjects (allowing for dropouts) was considered sufficiently large to detect significant and clinically important differences in physical performance measures between multiple and non-multiple faller groups (α = 0.05, β = 0.80 in two-sided tests). Thirty participants undertook the tests a second time 2 weeks after the initial assessment to determine the test/re-test reliability of the tests. The University of New South Wales Ethics Committee approved the study, and informed consent was obtained from subjects prior to their participation.

**Functional mobility tests**

**Sit-to-stand test**
The sit-to-stand test is used as a measure of lower limb strength [14] and is included in fall risk assessment scales [9, 15, 16]. For the sit-to-stand test with five repetitions (STS-5), subjects were asked to rise from a standard height (43 cm) chair without armrests, five times, as fast as possible with their arms folded. Subjects undertook the test barefoot and performance was measured in seconds, as the time from the initial seated position to the final seated position after completing five stands. The single sit-to-stand task (time from sitting to standing) (STS-1) was also evaluated as it has been used in assessment scales [17, 18] as a measure of functional mobility, balance and lower limb strength.

**Pick-up-weight test**
The ability to reach down and pick up an object from the floor has been included in several mobility assessment scales [7, 17]. In this study, a bag containing a 5 kg weight with handles that extended 50 cm above the floor was placed on the floor in front of the subject. They were asked to pick up the bag and place it on a table using one hand only. Performance was rated as either able or unable to complete the task.

**Half-turn test**
The ability to turn around in an efficient manner has been included in assessments of mobility and balance in older people [17, 19]. In this study, subjects were asked to take a few steps and then turn around to face the opposite direction. The number of steps taken to complete this 180° turn was counted.

**Alternate-step test**
The alternate-step test (AST) is a modified version of the Berg stool-stepping task [17]. It involves weight shifting and provides a measure of lateral stability. This test involved alternatively placing the entire left and right feet (shoes removed) as fast as possible onto a step that was 18 cm high and 40 cm deep. The time taken to complete eight steps, alternating between the left and right feet comprised the test measure.

**Mobility tests for predicting falls in older people**

**Six-metre-walk**
Slow gait speed is associated with an increased risk of falls [20, 21] and is a measure included in fall risk assessment scales [19, 22]. Subjects completed a six-metre-walk test (SMWT) measured in seconds along a corridor at their normal walking speed. A 2-m approach and a further 2 m beyond the measured 6-m distance ensured that walking speed was constant across the 6 m.

**Stair ascent and descent**
The inability to negotiate stairs is a marker of functional decline in older people [23] and many falls occur during this task [24]. In this study, the test stairs were indoors, had a handrail, were covered with linoleum and well lit. The subjects started the stair-ascent test at the bottom of eight steps (15 cm high, 27.5 cm deep). Subjects could use the handrail if preferred and a walking aid if they normally used one. Timing commenced for the stair-ascent test when the subject raised their foot off the ground to climb the first step and stopped when both feet were placed on the eighth step (which was a landing). After a brief rest, the subject was asked to descend the stairs. Timing was started when they raised their foot off the ground for the first step and stopped when they completed the last step. Time taken to complete the ascent and descent tests was recorded.

**Falls surveillance**
Falls were defined as ‘events that resulted in a person coming to rest unintentionally on the ground or other lower level, not as the result of a major intrinsic event or an overwhelming hazard’ [25]. Falls were monitored for 1 year with monthly fall calendars. If calendars were not returned, further contact was made by telephone.

**Statistical analysis**
Intra-class correlation coefficients (ICC1,1) and kappa statistics were calculated to evaluate the test/re-test reliability of the mobility tests. The relationships among the mobility tests were also examined with Pearson correlations and principal components analysis. Group t-tests were used to assess differences in mobility task performance for multiple fallers and non-multiple fallers based on the prospective falls data. Receiver-operated characteristics (ROC) curves were inspected to determine cut-off points for each continuously scaled mobility test that best discriminated between those who did and did not suffer multiple falls. Cut-off points for maximising the sensitivity and specificity for each mobility test were determined using the Youden Index [26]. Relative
A. Tiedemann et al.

risks, sensitivity, specificity and likelihood ratios were then calculated for the cut-off points and a logistic regression analysis was used to determine whether two or more mobility measures could be identified as significant and independent predictors of multiple falls. Finally, a $\chi^2$ test for trend was used to assess for any increased risk of falls with impairments in two or more tests. Analyses were performed using SPSS for Windows (Version 14.0, SPSS Inc., Chicago, IL, USA). Variables with right-skewed distributions were log$_{10}$ transformed in all analyses.

**Results**

The ICC$_{3,1}$ values indicated excellent reliability for the STS-5 (0.89, 95% CI = 0.79, 0.95), the AST (0.78, 95% CI = 0.59, 0.89), the half-turn test (0.75, 95% CI = 0.54, 0.87), the stair-ascent test (0.84, 95% CI = 0.69, 0.92) and the stair-descent test (0.86, 95% CI = 0.74, 0.93) [27]. The SMWT and the STS-1 displayed fair to good reliability (ICC$_{3,1}$ of 0.74 (95% CI = 0.52, 0.87) and 0.54 (95% CI = 0.23, 0.75) respectively). For the pick-up test, the kappa statistic could not be computed since all subjects in the reliability sub-study could complete the task on both occasions.

Appendix 1 shows the associations among the mobility tests (please see these supplementary data on the Journal website http://www.ageing.oxfordjournals.org/). There were significant correlations among all the mobility test scores. All the mobility tests were also significantly but weakly associated with age ($r = 0.12–0.34$, $P < 0.025$). The principal components analysis identified only one factor amongst the mobility tests that had an eigenvalue greater than 1, indicating that the tests were measuring a similar underlying construct. The absolute value loadings of the tests ranged from 0.72 to 0.93, and the factor explained 65% of the variance of the mobility test measures.

In all, 362 subjects (90%) completed the 12-month follow-up. Out of the non-completers, 4 died, 4 moved from the study area, 14 reported ill health and 16 withdrew consent. In the follow-up year 99 subjects suffered one fall (27.3%) and 80 subjects (22.1%) suffered two or more falls. Multiple fallers performed significantly worse than non-multiple fallers in the STS-5, SMWT, AST and the stair-descent tests (Table 1). Multiple fallers and non-multiple fallers showed no evidence of a difference with respect to age ($t_{360} = 1.39$, $P = 0.17$), and the proportion of men and women in the two faller groups were similar ($\chi^2 = 0.01$, $df = 1$, $P = 0.94$).

The cut-off points for each test and associated statistics are shown in Table 2. The AST had the highest relative risk (RR) for prediction of multiple fallers (2.3) and displayed good sensitivity (69%) but only moderate specificity (56%). The other tests had RRs ranging from 1.3 to 2.0 and sensitivity and specificity scores ranging from 11 to 78%, and 28 to 93% respectively. All the tests displayed likelihood ratios which were greater than 1, and indicated that poor performance in the tests was weakly associated with a greater risk of being a multiple faller.

The logistic regression analysis revealed that the AST was the only variable which was significantly associated with falls in a model containing the four mobility measures found to be significant predictors of falls in bi-variate analyses (OR = 2.81, 95% CI = 1.62, 4.86). However, other models including only the STS-5 or the SMWT resulted in similar predictive ability (OR = 2.40, 95% CI = 1.43, 3.03 and OR = 2.10, 95% CI = 1.27, 3.48).

The $\chi^2$ test for trend analysis revealed poor performance in two mobility tests increased the risk of multiple falls more than poor performance in one test alone $- \chi^2 = 16.46$, $df' = 4$, $P < 0.001$. This is illustrated by an OR of 3.66 (95% CI = 1.44, 9.27) for poor performance in any two of the four predictive mobility tests (sensitivity = 67%, specificity = 65%), compared with an OR of 1.61 (95% CI = 0.62, 4.16) for poor performance in one test alone (sensitivity = 60%, specificity = 52%). However, there was little additional increased risk of multiple falls for poor performances in three or all four tests (ORs = 4.85

---

**Table 1.** Comparison between non-multiple fallers and multiple fallers in performance of mobility tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Non-multiple fallers</th>
<th>Multiple fallers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean</td>
</tr>
<tr>
<td>Sit-to-stand once (s)$^{a,b}$</td>
<td>170</td>
<td>1.0</td>
</tr>
<tr>
<td>Sit-to-stand five times (s)</td>
<td>282</td>
<td>12.5</td>
</tr>
<tr>
<td>Pick-up-weight test (number (%) able to complete)$^{c}$</td>
<td>282</td>
<td>262 (93)</td>
</tr>
<tr>
<td>Half-turn test steps—(Median and inter-quartile range)$^{d}$</td>
<td>282</td>
<td>4.5</td>
</tr>
<tr>
<td>Alternate-step test (s)$^{a,b}$</td>
<td>265</td>
<td>10.8</td>
</tr>
<tr>
<td>Six-metre walk (s)$^{a,b}$</td>
<td>282</td>
<td>5.8</td>
</tr>
<tr>
<td>Stair ascent (s)$^{a,b}$</td>
<td>282</td>
<td>5.5</td>
</tr>
<tr>
<td>Stair descent (s)$^{a,b}$</td>
<td>282</td>
<td>5.7</td>
</tr>
</tbody>
</table>

---

$^{a}$ Lower group numbers due to introduction of the tests after commencement of the trial.
$^{b}$ Raw data presented, but analyses performed on log$_{10}$ transformed data.
$^{c}$ Mann–Whitney U test.
$^{d}$ Pearson $\chi^2$ test.
AST and the SMWT were identified as the best tests. Using this weighting system, the STS-5, the stair-descent test, when analysed, had poor performances in two mobility tests. This suggests that impairments in two tests, e.g., reduced ability to rise from a chair and slow walking speed, had a cumulative effect on fall risk. However, as subjects’ performances across the tests were significantly correlated, poor performances in more than two tests resulted in little additional increased risk of falls. This indicates that one, or at most two, valid tests could serve as a useful screen for referral and further assessment [28]. Additional tests, however, would assist with the tailoring of exercise interventions to specific impairments, such as difficulty in climbing stairs and for evaluating such interventions. All the tests included in this study, with the exceptions of the pick-up-weight and the sit-to-stand-once tests had acceptable test/re-test reliability, and thus seem appropriate for use in this regard.

A final factor in the evaluation of tests useful for screening purposes is an assessment of feasibility in clinical practice. Of the tests assessed, the sit-to-stand and the half-turn tests are the most feasible as they require no equipment, and little space for test administration. The AST and the pick-up-weight test are also feasible but require additional simple equipment. The six-metre walk, the stair ascent and descent tests may need to be performed in busy corridors and stair wells, making them less feasible.

Table 2. Cut-off points for mobility tests and associated sensitivity, specificity and relative risk (RR) statistics with respect to non-multiple faller—multiple faller comparisons

<table>
<thead>
<tr>
<th>Test</th>
<th>Criterion</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Relative risk (95% CI)</th>
<th>Likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sit-to-stand once</td>
<td>≥ 1s</td>
<td>0.49 (0.34, 0.64)</td>
<td>0.58 (0.50, 0.66)</td>
<td>1.3 (0.8, 2.1)</td>
<td>1.17</td>
</tr>
<tr>
<td>Sit-to-stand five times</td>
<td>≥ 12s</td>
<td>0.66 (0.55, 0.76)</td>
<td>0.55 (0.49, 0.61)</td>
<td>2.0 (1.3, 3.0)</td>
<td>1.47</td>
</tr>
<tr>
<td>Pick-up-weight test</td>
<td>Unable</td>
<td>0.11 (0.06, 0.21)</td>
<td>0.93 (0.89, 0.96)</td>
<td>1.5 (0.8, 2.6)</td>
<td>1.57</td>
</tr>
<tr>
<td>Half-turn test</td>
<td>≥ 4 steps</td>
<td>0.78 (0.67, 0.86)</td>
<td>0.28 (0.23, 0.34)</td>
<td>1.3 (0.8, 2.0)</td>
<td>1.08</td>
</tr>
<tr>
<td>Alternate-step test</td>
<td>≥ 10s</td>
<td>0.69 (0.57, 0.79)</td>
<td>0.56 (0.50, 0.62)</td>
<td>2.3 (1.4, 3.5)</td>
<td>1.57</td>
</tr>
<tr>
<td>Six-metre walk</td>
<td>≥ 6s</td>
<td>0.50 (0.39, 0.61)</td>
<td>0.68 (0.62, 0.73)</td>
<td>1.8 (1.2, 2.6)</td>
<td>1.56</td>
</tr>
<tr>
<td>Stair ascent</td>
<td>≥ 5s</td>
<td>0.54 (0.42, 0.65)</td>
<td>0.58 (0.52, 0.64)</td>
<td>1.4 (1.0, 2.1)</td>
<td>1.29</td>
</tr>
<tr>
<td>Stair descent</td>
<td>≥ 5s</td>
<td>0.63 (0.51, 0.73)</td>
<td>0.55 (0.49, 0.61)</td>
<td>1.7 (1.2, 2.6)</td>
<td>1.40</td>
</tr>
</tbody>
</table>

Table 3. Validity, reliability and feasibility weightings for the mobility tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Validity</th>
<th>Reliability</th>
<th>Feasibility</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sit-to-stand five times</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>Alternate-step test</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>Six-metre walk</td>
<td>10</td>
<td>4</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>Stair ascent</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Stair ascent</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Half-turn test</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Sit-to-stand once</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Pick-up-weight test</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Validity (external validity—ability to discriminate between multiple and non-multiple fallers):
10—significant predictor of fall risk when coded as both a continuous and discrete variable.
5—borderline significant predictor of fall risk when coded as a discrete variable ($P = 0.05$).
0—not a significant predictor of fall risk.

Reliability (test/re-test reliability):
5—ICC > 0.8; 4—ICC > 0.7; 3—ICC > 0.6; 2—ICC > 0.5; 1 for ICC < 0.5 or not calculable.

Feasibility:
5—no equipment required, can be administered in small treatment room.
4—minimal equipment required, can be administered in small treatment room.
3—extended administration space, such as a corridor required.
0—requires assessment on a flight of stairs, impractical in many clinical settings.

(95% CI = 1.86, 12.65) and 4.31 (95% CI = 1.79, 10.40) respectively.

Table 3 presents a weighting system based on validity, reliability and feasibility for determining the optimal tests. Validity was given double the weighting of feasibility, as ability to predict older people at risk of falls is paramount in this context. Using this weighting system, the STS-5, the AST and the SMWT were identified as the best tests.

Discussion

This study assessed the relative ability of eight mobility tests to predict multiple fallers in a large, representative sample of older people. Multiple fallers performed worse than non-multiple fallers in four of the tests: the STS-5, the AST, the SMWT and the stair-descent test, when analysed as continuous variables. When dichotomised, the AST was the best test for discriminating between the faller groups.

An AST cut-off point of 10 s was associated with a 130% increased risk, with 69% sensitivity and 56% specificity with respect to identifying multiple fallers. At identified cut-off points, the STS-5 (12 s), the SMWT (6 s), the stair-descent test (5 s) and the stair-ascent test (5 s) could also significantly predict subjects who suffered multiple falls with sensitivities and specificities above 50%.

A single mobility test measure (the AST) provided the best discrimination between multiple and non-multiple fallers in a logistic regression analysis. With this test in the regression model, the remaining test measures could add minimal additional discrimination between the falls outcome groups due to multi-collinearity among the mobility tests. This was supported by a principal components analysis that identified only one factor, indicating that to a large extent the tests were measuring the same underlying mobility construct.

There was an increased risk of multiple falls if subjects had poor performances in two mobility tests. This suggests that impairments in two tests, e.g., reduced ability to rise from a chair and slow walking speed, had a cumulative effect on fall risk. However, as subjects’ performances across the tests were significantly correlated, poor performances in more than two tests resulted in little additional increased risk of falls. This indicates that one, or at most two, valid tests could serve as a useful screen for referral and further assessment [28]. Additional tests, however, would assist with the tailoring of exercise interventions to specific impairments, such as difficulty in climbing stairs and for evaluating such interventions. All the tests included in this study, with the exceptions of the pick-up-weight and the sit-to-stand-once tests had acceptable test/re-test reliability, and thus seem appropriate for use in this regard.

A final factor in the evaluation of tests useful for screening purposes is an assessment of feasibility in clinical practice. Of the tests assessed, the sit-to-stand and the half-turn tests are the most feasible as they require no equipment, and little space for test administration. The AST and the pick-up-weight test are also feasible but require additional simple equipment. The six-metre walk, the stair ascent and descent tests may need to be performed in busy corridors and stair wells, making them less feasible.
A. Tiedemann et al.

It is acknowledged that as the participants comprised relatively healthy older people, the findings are unlikely to generalise to frail older people. Secondly, although we measured the component parts, we did not include the timed up-and-go test in our battery. This would have made for useful comparisons as the timed up-and-go test is recommended by the American and British Geriatrics Societies as the screening test for falls, despite insufficient validation [28].

The identification of a single, simple diagnostic test with excellent predictive accuracy would be a boon for the assessment of falls risk in older people. None of the assessed tests reached this standard. However, it is unlikely that any single mobility test will be shown to have excellent predictive accuracy, as it is known that the causes of falls are multifactorial with several unrelated to mobility. For example, poor vision, cardiovascular conditions, cognitive impairment and risk taking can lead to falls independent of mobility limitations. Further, predictive accuracy is limited by the inherently problematic nature of the outcome variable (i.e. falls), which relies on self-report. The identification of a single mobility measure for the accurate prediction of falls is therefore unrealistic and such tests should be used only as initial screens for identifying older people in need of further assessment. The AST, the STS-5 and the SMWT are the best tests in this regard, as they meet acceptable standards of validity, reliability and feasibility.

Key points
- The sit-to-stand, pick-up-weight, half turn, alternate step, six-metre walk, stair ascent and stair-descent tests are feasible and reliable mobility tests for community-dwelling older people.
- The tests demonstrated poor to fair sensitivity and specificity in identifying older people at risk of multiple falls. These tests should therefore only be used as initial screens for identifying older people in need of further assessment.
- The mobility tests appear to measure a similar underlying ‘mobility’ construct. In consequence, one or at most two tests are required in screening older people at risk of falls.
- Based on assessments of feasibility, reliability and predictive validity for falls, the alternate step, the STS-5 and the six-metre-walk tests were the best tests.

Acknowledgements

The National Health and Medical Research Council [Population Health Capacity Building Grant in Injury Prevention, Trauma and Rehabilitation and Health Research Partnership Grant: Prevention of Older People’s Injuries (POPI)], MBF Australia, and the Vincent Fairfax Family Foundation supported this project. The authors would like to acknowledge Dr Susan Ogle, Head of the Aged Care and Rehabilitation Department for the use of the testing facilities at the Royal North Shore Hospital.

Supplementary data

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

References

Mobility tests for predicting falls in older people


Received 28 January 2007; accepted in revised form 5 February 2008