A Comparative Study of the Use of Four Fall Risk Assessment Tools on Acute Medical Wards

Michael Vassallo, FRCP, PhD,* Rachel Stockdale, MRCP (UK), ‡ Jagdish C. Sharma, FRCP,‡
Roger Briggs, FRCP,† and Stephen Allen, FRCP§

OBJECTIVES: To compare the effectiveness of four falls risk assessment tools (STRATIFY, Downton, Tullamore, and Tinetti) by using them simultaneously in the same environment.

DESIGN: Prospective, open, observational study.

SETTING: Two acute medical wards admitting predominantly older patients.

PARTICIPANTS: One hundred thirty-five patients, 86 female, mean age ± standard deviation 83.8 ± 8.01 (range 56–100).

MEASUREMENTS: A single clinician prospectively completed the four falls risk assessment tools. The extent of completion and time to complete each tool was recorded. Patients were followed until discharge, noting the occurrence of falls. The sensitivity, specificity, negative predictive accuracy, positive predictive accuracy, and total predictive accuracy were calculated.

RESULTS: The number of patients that the STRATIFY correctly identified (n = 90) was significantly higher than the Downton (n = 46; P < .001), Tullamore (n = 66; P = .005), or Tinetti (n = 52; P < .001) tools, but the STRATIFY had the poorest sensitivity (68.2%). The STRATIFY was also the only tool that could be fully completed in all patients (n = 135), compared with the Downton (n = 130; P = .06), Tullamore (n = 130; P = .06), and Tinetti (n = 17; P < .001). The time required to complete the STRATIFY tool (average 3.85 minutes) was significantly less than for the Downton (6.34 minutes; P < .001), Tinetti (7.4 minutes; P < .001), and Tullamore (6.25 minutes; P < .001). The Kaplan-Meier test showed that the STRATIFY (log rank P = .001) and Tullamore tools (log rank P < .001) were effective at predicting falls over the first week of admission. The Downton (log rank P = .46) and Tinetti tools (log rank P = .41) did not demonstrate this characteristic.

CONCLUSION: Significant differences were identified in the performance and complexity between the four risk assessment tools studied. The STRATIFY tool was the shortest and easiest to complete and had the highest predictive value but the lowest sensitivity. J Am Geriatr Soc 53:1034–1038, 2005.

Key words: falls; risk assessment; comparative study

Because there is some evidence that falls in hospital can be reduced,1–3 it is important to identify high-risk patients likely to benefit from expensive multidisciplinary interventions.4 Several fall risk factors have been identified, and some of them have been compiled into fall risk assessment tools.5,6 Such tools are based on the premise that the higher the number of risk factors, the higher the risk of falling.7

Although a number of tools have been used to identify fall risk in hospitalized patients,5,6 not all have been validated.9 Several of those that have been validated have high accuracy, but when tested outside the specific setting in which they were originally validated, the predictive accuracy is not reproduced.10 This may have occurred because of different patient and staff characteristics, as well as operational differences between the environments. There is considerable overlap between the characteristics used to compile the tools, raising the question of whether fall risk assessment tools are indeed any different. A comparative study of four assessment tools was therefore conducted to determine whether there are any differences in performance of the tools in the same environment. The chosen fall risk tools were the STRATIFY,10 the Tullamore,11 a Tinetti-based assessment,11 and the Downton.12 They were chosen because staff were familiar with their use, they had previously been used on elderly people in various hospital settings to predict the risk of falls,1–8,10,11,13 and they are still being used widely in many hospitals.

By studying four fall risk assessment tools simultaneously with the same investigators, under the same ward
conditions, it was hoped to determine whether there are real differences in effectiveness between fall risk assessment tools and whether more complex tools are any better than simple tools at identifying patients who fall on medical wards.

METHODS
The study was conducted in two medical wards admitting predominantly elderly patients. They were admitted for treatment of a wide range of medical conditions. Approval was obtained from the North Nottinghamshire ethics committee. One hundred thirty-five consecutive patients were studied. None declined to participate. On admission, patients had medical and nursing assessments. A single clinician prospectively conducted the medical assessment, which consisted of measurements of vision, depression, mobility, and a medication review. The nursing assessment included information on agitation, need of frequent toileting, activities of daily living, and postural blood pressure. The clinician completing the medical assessment completed the risk assessment tool. Fall prevention measures were not dependent on the score obtained, but measures were taken to try to correct any individual fall risk factors identified using the various tools. For example, postural hypotension would be treated even though the patient might be at low overall risk of falls.

Information was collected on patients’ age, sex, history of falls, and medications on admission. Patients had a physical examination noting the presence of impaired vision, hearing loss, lower limb abnormalities, gait disturbance, back extension, postural hypotension, and confusion. Patients were deemed to have impaired vision if they were registered blind or partially sighted or were unable to see less than 6/60 on a Snellen chart using glasses, if appropriate. Hearing impairment was defined as the inability to follow a conversation with or without using a hearing aid. A limb was considered abnormal if there was any evidence of weakness (Medical Research Council criteria grade 4/5 or less), neuropathy, amputation, joint abnormality excluding minor osteoarthritic changes, or any condition judged to interfere with normal gait such as cellulitis or a deep vein thrombosis. A patient’s gait was assessed using the Get Up and Go Test.14 On this basis, patients were classified into four groups: normal, safe (with or without using aids), unsafe (with or without using aids), and unable, if the patient was bedridden. Back extension was studied after testing the patient’s mobility and was recorded with the patient standing as being able or unable to perform the maneuver. Patients were considered to be confused if they scored less than 7 of 10 on the Hodkinson Abbreviated Mental Test score.15

The Downton Fall risk tool12 was compiled based on a history of falls, medications (tranquilizers/sedatives, diuretics, antihypertensives excluding diuretics, antiparkinsonian drugs, and antidepressants), sensory deficits (visual impairment, hearing impairment), limb abnormalities (such as hemiparesis), confusion, and unsafe gait (with or without aids). Each of these factors scored a point; scores of 3 or above identify patients at risk.

STRATIFY consists of five factors, each found to be independently associated with falling.10 These factors are presenting with a fall or having a fall on the ward, the presence of agitation, visual impairment, need for frequent toileting, and impaired ability to transfer and walk. Scores of 2 or more were considered to be high risk.

The Tinetti fall risk index is based on number of chronic disabilities.7 The higher the number of chronic disabilities, the higher the likelihood of having recurrent falls. The nine risk factors included in the fall risk index are mobility score, morale score, mental status score, distance vision, hearing, postural blood pressure drop, back examination, medications on admission, and admission activity of daily living score. To simplify the tool, the Geriatric Depression Scale score, Get Up and Go test, and Abbreviated Mental Test score were used instead of the Philadelphia Morale score, gait and balance assessment, and Mini-Mental State Examination score, respectively. These modifications were validated for use16 (unpublished data). The subjects’ fall risk score was the number of index risk factors present. Scores of 0 to 3 were considered low, 4 to 6 was medium, and 7 to 9 was high risk. For the purposes of analysis, medium and high risk were considered together.

The Tullamore tool11 assesses sex, age, gait, sensory deficits, falls history, medication, medical history and mobility under various subheadings. Patients are classified as low (score 3–8), medium (score 9–12) or high (score 13). Medium- and high-risk scores were considered together.

For all the tools, the cutoff point from low to higher risk was that suggested by the respective authors. Patients were followed up to the point of discharge from the ward. Nursing staff kept a record of falls as they occurred on the wards in a falls diary. The clinician completing the tool was blinded to the occurrence of falls that occurred after tool completion. Patients who fell at least once were classified as fallers. Other outcomes assessed were the number of risk assessments that could be completed in their entirety on initial assessment and how long it took to complete the falls risk assessment. The time was calculated by estimating the time taken for each individual assessment. The total for the tool was then the total of all the individual components required to complete the tool.

STATISTICS
The sensitivity, specificity, and total predictive accuracy of the tools were calculated. Sensitivity was defined as the total number of fallers correctly identified as high risk. Specificity was defined as total number of nonfallers correctly defined as low risk. The total predictive accuracy was the total number of patients correctly identified expressed as a percentage. The positive predictive value was defined as the number of high-risk patients who went on to fall whereas the negative predictive value was the number of low-risk patients who did not fall. Results were expressed as a percentage. Fishers exact probability test was used to compare the accuracy of the various risk tools by comparing the numbers of patients correctly identified by the various tools to the best performing tool. Data was collected on 135 patients. Assessment items that could not be completed were identified and recorded. Because we aimed to evaluate the practical utility of each of the tools, all items were included in calculating the total score. Incomplete items received a
score of zero so that the total score for each tool was derived from the number of positively scored items.

The Kaplan-Meier hazard statistic was used to assess the likelihood of falls in high- and low-risk patients for each of the tools for the first week of patient stay using all 135 patients. Significance was expressed using the Log rank test. The number of falls was censored in daily time intervals for the first week in both the high- and low-risk category for each of the tools.

RESULTS

One hundred thirty-five patients were studied: 86 female, mean age ± standard deviation 83.8 ± 8.01 (range 56–100). Almost all patients had an acute illness of varying severity (e.g., respiratory tract infection, heart failure, stroke, urinary tract infection) on a background of chronic disease (e.g., arthritis or dementia). The mean length of stay was 14.6 ± 7.5 days (range 6–22 days). Twenty-two fallers, of whom six had recurrent falls, contributing 29 falls in total, were identified. The performance of the various tools is shown in Table 1. The STRATIFY had the highest total predictive accuracy but the lowest sensitivity. The number of patients that the STRATIFY correctly identified (number of high-risk patients who fell and low-risk patients who did not fall) was significantly higher than the Downton (P < .001), Tullamore (P = .005), and Tinetti (P < .001). In view of the low sensitivity, a separate analysis with 1 or above as the cutoff point for high risk was performed. This change gave a sensitivity of 86% and specificity of 25%. The number of correctly identified patients was 48 (35%). This was significantly inferior to the Tullamore (P = .03), which now had the highest number of correctly identified patients.

It was possible to complete the STRATIFY tool for all patients evaluated. None of the other tools could be completed in their entirety for all patients (Table 2). The STRATIFY performed significantly better than the Tinetti (P < .001), which could be completed for only 17 patients. Not all the Tinetti items could be completed because of inability to perform postural blood pressure measures (n = 75) or the Geriatric Depression Scale (n = 62) because of severe cognitive function, although it was still possible to classify 70% of patients as medium- to high-risk for falls using the Tinetti tool. Completion of the STRATIFY did not differ from the Downton (P = .06) and Tullamore tools (P = .06). The Downton and Tullamore were not completed in five subjects because of an inability to complete a cognitive assessment. It was still possible to classify the respective patients into a high-risk category. The time required to fill in the STRATIFY was significantly less than that for the Downton (P < .001), Tinetti (P < .001), and Tullamore (P < .001).

The predictive value of the tools to identify fallers over the first week was analyzed using the Kaplan-Meier hazard test. The STRATIFY (log rank P = .002) and Tullamore (log rank P < .001) were able to identify fallers from the time of admission and throughout the first week of patient stay (Figure 1). The Downton (log rank P = .46) and Tinetti (log rank P = .42) did not demonstrate a similar ability.

DISCUSSION

It is well recognized that the performance of any given fall risk assessment tool varies when used in different settings. This may result from differences in patient, staffing, and environmental characteristics. By studying and comparing the performance of the various tools under the same conditions in the same environment, this study attempted to identify any differences in the effectiveness of these tools. Although there is considerable overlap between the

| Table 1. Characteristics of Tools When Identifying Fallers |
|----------------|----------------|----------------|----------------|
| Characteristic | Downton (n = 135) | STRATIFY (Medium/high risk) (n = 135) | Tullamore (Medium/high risk) (n = 135) |
| Sensitivity    | 81.8            | 68.2            | 90.9            |
| Specificity    | 24.7            | 66.4            | 40.7            |
| Positive predictive value | 17.5          | 28.3            | 22.9            |
| Negative predictive value | 87.5          | 91.5            | 95.8            |
| Patients correctly identified, n* | 46            | 90              | 66              |
| Total predictive accuracy, % | 34.1           | 66.6            | 48.8            |

* STRATIFY vs Downton, P < .001; Tullamore, P = .005; Tinetti, P < .001.

| Table 2. Number of Completed Risk Assessment Tools and Time to Complete |
|----------------|----------------|----------------|----------------|
| Other Score Characteristic | Downton (n = 135) | STRATIFY (Medium/high risk) (n = 135) | Tullamore (Medium/high risk) (n = 135) |
| Time to complete, minutes, mean ± standard deviation | 6.34 ± 2.62 | 3.85 ± 1.67 | 6.25 ± 2.56 |
| Number fully completed | 130 | 135 | 130 |

* STRATIFY vs Downton, P < .001; Tullamore, P < .001.

† STRATIFY vs Tinetti, P < .001.
characteristics compiling the various tools, this study identified significant differences in their performance.

The total predictive accuracy of the various tools was low, the highest being the STRATIFY at 66.6%. This is principally because of low specificity. Specificity refers to patients who do not fall having been correctly identified as at low risk of falling. However, the hallmark of effective fall prevention measures and high-quality care is that high-risk patients are prevented from falling. A low specificity is expected in an environment in which patients are prevented from falling. In addition, not all patients identified as being at high risk will fall even if left on their own, further reducing a tool’s total predictive accuracy. The ideal tool, with high sensitivity and specificity, is difficult to develop, and the most important measure of a fall risk assessment tool is arguably its sensitivity. A possible way of improving fall risk assessment is to focus on items in existing tools that improved their sensitivity, such as a history of falls, confusion, and an unsafe gait. The STRATIFY, despite having the highest total predictive accuracy, had the lowest sensitivity because it failed to identify the highest number of fallers as being at high risk. This is an important weakness of the tool. Changing the cutoff point of the tool to 1 improved its sensitivity by increasing the number of fallers identified as high risk but reduced the predictive accuracy.

The tools differed in complexity, as reflected in the significant differences in the time required to fill them out and the inability to complete some of the tools in their entirety. It is likely that the low accuracy of some tools like the Tinetti resulted from the missing data. Despite this, it was still possible to categorize the majority of patients into the medium- to high-risk group. The analyses included all patients, regardless of whether the tools were completed in their entirety because it was desired to study the utility of it in real life, where one needs to decide on the outcome of a risk assessment regardless of whether a tool is completed. This is important because fall risk assessment needs to be accurate, simple, and not time consuming to be implemented effectively on wards without adding a considerable burden on already hard-pressed staff.

This study has a number of limitations. The fall risk assessment was done only once; therefore a change in patient condition could explain the low predictive value of the tools. The results obtained may not be reproducible on other units because of differing staff and patient characteristics, including a different sex mix. Another limitation is that it did not include all available fall risk assessment tools. Some, such as the Morse Fall Scale, are widely used in the United States, but significant differences were identified between the four risk assessment tools studied, with STRATIFY having the best predictive accuracy but the lowest sensitivity.

REFERENCES


