Fall risk-assessment tools compared with clinical judgment: an evaluation in a rehabilitation ward

MICHAEL VASSALLO, LYNN POYNTER, JAGDISH C. SHARMA, JOSEPH KWAN, STEPHEN C. ALLEN
Royal Bournemouth Hospital, Medicine Castle Lane East Bournemouth, Dorset BH7 7DW, UK

Abstract

Objectives: to compare the use of two falls risk-identification tools (Downton and STRATIFY) with clinical judgment (based upon the observation of wandering behaviour) in predicting falls of medically stable patients in a rehabilitation ward for older people.

Methods: in a prospective observational study, with blinded end-point evaluation, 200 patients admitted to a geriatric rehabilitation hospital had a STRATIFY and Downton Fall Risk assessment and were observed for wandering behaviour.

Results: wandering had a predictive accuracy of 78%. A total of 157/200 were identified correctly compared to 100/200 using the Downton score ($P<0.0001$ 95%, CI 0.18–0.42), or 93/200 using STRATIFY ($P<0.0001$; 95% CI 0.15–0.37). The Downton and STRATIFY tools demonstrated predictive accuracies of 50% and 46.5%, respectively, with no statistical significance between the two ($P = 0.55$; 95% CI 0.77–1.71). Sensitivity for predicting falls using wandering was 43.1% (22/51). This was significantly worse than Downton 92.2% (47/51: $P<0.001$) and STRATIFY 82.3% (42/51: $P<0.001$).

Conclusions: this study showed that clinical observation had a higher accuracy than two used falls risk-assessment tools. However it was significantly less sensitive implying that fewer patients who fell were correctly identified as being at risk.

Keywords: falls, hospital, risk-assessment tools, clinical judgment, wandering, elderly

Introduction

It is often stated that the risk of falls is proportional to the number of risk factors [1]. On this premise, a considerable number of risk-assessment tools have been developed [2–5]. Despite claims of high accuracy, these tools are often found to be disappointing when used outside the context of their original validation [6–9]. Scores that are effective in the circumstances in which they have been developed are not found to be effective when used outside the place of development even when the clinical setting and case mix is very similar [10–12]. There are several possible reasons for such variations in accuracy. There might be poor validation of the tool in the first place [13]. However, even when tools are properly validated, the characteristics of patient populations and staff vary in different hospitals despite having similar roles, thus explaining such variance in performance [6]. For example, staff using the tool might interpret the same observation differently. Since current tools are found to be inadequate, new ones continue to be developed [14]. There are even calls for the development of more diagnosis-specific tools [15].

The current limitations associated with falls risk-assessment tools stimulate the need to rethink how they can contribute to managing falls risk [16]. In hospital, a small number of risk factors consistently emerge as important predisposing factors for falls [17]. These include gait instability, agitated confusion, urinary incontinence, urinary frequency, falls history and drugs, particularly sedatives and hypnotics [6]. It has been suggested that these common risk factors for falls should be identified and acted upon when formulating a care plan for all patients who have those factors [6]. This contrasts with the alternative model seeking the development of further falls risk-assessment tools for use in specific environments and targeting only high-risk patients for intervention [15].

In this study, we aimed to assess the risk of falls in patients in a rehabilitation ward using two established falls risk-assessment tools, and a clinical judgment of risk based on the observation of wandering behaviour. This approach has its origin partly in reports that have suggested that an intuitive assessment of falls risk by nursing staff can be as effective as using falls risk-assessment tools [18–20]. Formal falls risk assessment is a time-consuming process and requires considerable resources and commitment to complete [21]. If clinical judgment in a hospital environment is just as effective as risk-assessment tools a case can be made to re-evaluate the way falls risk management is conducted.
Methods

The study was carried out in one rehabilitation ward of a rehabilitation hospital admitting elderly patients. The patients were admitted for rehabilitation after treatment for a wide range of medical conditions. Approval was obtained from the North Nottinghamshire Ethics Committee. We studied 200 consecutive patients so the sample was quasi-random. None declined to participate. On admission, patients had a medical and nursing assessment. A single clinician conducted the medical assessment prospectively. The clinician conducting the medical assessment completed the risk-assessment tools.

Information was collected on the patients’ age, gender, past 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 and confusion. Patients were deemed to have impaired vision if they were registered blind or partially sighted or were unable to see better 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 (MRC 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 by performing the Get Up and Go Test [22]. 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 bed ridden. Patients were considered to be confused if they scored less than 7/10 on the Abbreviated Mental Test score [23].

Clinical judgment of falls risk was based on the observation of wandering behaviour, which was considered to predispose to falls. Wandering behaviour was defined for this study as a tendency to move about in either a seemingly aimless or disoriented fashion or in pursuit of an indefinable aim or goal [24, 25]. For this purpose, a descriptive topology was used. This included nine items, one of which had to be identified on more than one occasion within the first 48 h of admission. These were: checking, pottering, aimless walking, walking with inappropriate purpose, walking with appropriate purpose but inappropriate frequency, excessive activity, nighttime walking, attempts to leave the hospital and being brought back to hospital [26]. The assessment made by the clinician was based on the observation of any one of the above nine defined behaviours by the clinician himself and/or nurses who observed the patient out-of-hours.

The Downton Fall risk tool [4] was compiled on the basis of a past history of falls, medications (tranquillisers/sedatives, diuretics, anti-hypertensives excluding diuretics, anti-parkinsonian drugs, and antidepressants), sensory deficits (visual impairment, hearing impairment), limb abnormalities (such as hemiparesis), confusion and an unsafe gait (with or without aids). Each one of these factors score a point, and scores of three or above identified patients at risk.

STRATIFY consists of five factors, each found to be independently associated with falling [3]. These factors were: presenting with a fall or having a fall in the ward; the presence of agitation; visual impairment; need for frequent visits to the toilet and an impaired ability to transfer and walk. Scores of two or more were considered to be high-risk.

For all the tools, the cut-off point from low to a higher risk was that suggested by the respective authors. Patients were followed-up to the point of discharge from the ward. A record of falls was kept by nursing staff as they occurred, using a falls diary. Keeping the falls diary and filling accident forms were statutory requirements. 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 and those falling more than once as recurrent fallers.

Statistical analysis

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 non-fallers 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. The negative predictive value was the number of low-risk patients who did not fall. Results were expressed as a percentage. Fisher’s exact probability test was used to test categorical datasets.

Results

We studied 200 patients (mean age 80.9 years, 123 female). We identified 51 fallers and 17 recurrent fallers. The total predictive accuracy of observation of wandering behaviour was the highest at 78% compared to the Downton and STRATIFY scores (Table 1). Using wandering, 157/200 were identified correctly compared to 100/200 using the Downton score (P < 0.0001; 95% CI 0.18–0.42), or 93/200 using STRATIFY (P < 0.0001; 95% CI 0.15–0.37). No significant differences were identified between the two risk-assessment tools (P = 0.55; 95% CI 0.77–1.71).

Sensitivity for predicting falls using wandering was 43.1% (22/51). This was significantly worse when compared to the Downton 92.2% (47/51; P < 0.001) and STRATIFY 82.3% (42/51; P < 0.001). Conversely, wandering behaviour was found to have a significantly higher specificity compared to the two risk-assessment tools. This contributed significantly to the higher overall accuracy. Concordance between fallers identified as wanderers and identified as high-risk by the score was high for both the Downton (100% 22/22) and STRATIFY (86.4%; 19/22).

Similar observations were noted in relation to recurrent falls (Table 2) Wandering behaviour was most accurate (83.5%) at predicting recurrent fallers. Using wandering, 167/200 were identified correctly compared to 74/200 using...
Fall risk-assessment tools in a rehabilitation ward

Table 1. Characteristics of tools and wandering behaviour when identifying all fallers

<table>
<thead>
<tr>
<th></th>
<th>Downton</th>
<th>STRATIFY</th>
<th>Wandering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>92.2 (47/51), 0.82–0.97</td>
<td>82.3 (42/51), 0.69–0.90</td>
<td>43.1 (22/51), 0.30–0.56</td>
</tr>
<tr>
<td>CI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>35.8 (53/149), 0.28–0.43</td>
<td>34.2 (51/149), 0.27–0.42</td>
<td>90.6 (135/149), 0.84–0.94</td>
</tr>
<tr>
<td>Positive predictive value (%)</td>
<td>33.1 (47/143), 0.25–0.41</td>
<td>30.0 (42/140), 0.23–0.38</td>
<td>61.1 (22/36), 0.44–0.75</td>
</tr>
<tr>
<td>No. of patients correctly identified</td>
<td>93</td>
<td>93</td>
<td>157</td>
</tr>
<tr>
<td>Total predictive accuracy (%)</td>
<td>50.0 (100/200), 0.43–0.56</td>
<td>46.5 (93/200), 0.39–0.53</td>
<td>78.0 (157/200), 0.72–0.83</td>
</tr>
</tbody>
</table>

Table 2. Characteristics of tools and wandering when identifying recurrent fallers

<table>
<thead>
<tr>
<th></th>
<th>Downton</th>
<th>STRATIFY</th>
<th>Wandering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>100 (17/17), 0.81–0.99</td>
<td>94.1 (16/17), 0.76–0.99</td>
<td>58.8 (10/17), 0.35–0.78</td>
</tr>
<tr>
<td>CI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>31.1 (57/183), 0.24–0.38</td>
<td>32.2 (59/183), 0.26–0.39</td>
<td>85.8 (157/183), 0.79–0.90</td>
</tr>
<tr>
<td>Positive predictive value (%)</td>
<td>11.8 (17/143), 0.07–0.18</td>
<td>11.4 (16/140), 0.07–0.17</td>
<td>27.7 (10/36), 0.16–0.44</td>
</tr>
<tr>
<td>Negative predictive value (%)</td>
<td>100 (57/57), 0.93–0.99</td>
<td>98.3 (59/60), 0.91–0.99</td>
<td>95.7 (157/164), 0.91–0.97</td>
</tr>
<tr>
<td>No. of patients correctly identified</td>
<td>74</td>
<td>75</td>
<td>167</td>
</tr>
<tr>
<td>Total predictive accuracy (%)</td>
<td>37.0 (74/200), 0.31–0.43</td>
<td>37.5 (75/200), 0.31–0.44</td>
<td>83.5 (167/200), 0.77–0.87</td>
</tr>
</tbody>
</table>

The Downton score ($P<0.0001$; 95% CI 0.07–0.19) or 75/200 using STRATIFY ($P<0.0001$; 95% CI 0.07–0.19). No significant differences were identified between the two risk-assessment tools ($P=1.0$; 95% CI 0.64–1.49). Wandering was found to have a significantly higher specificity compared to the two risk-assessment tools similarly contributing significantly to the higher overall accuracy.

The predictive value of the tools and of wandering behaviour to identify fallers and recurrent fallers over the first 21 days of patient stay was analysed using the Kaplan–Meier hazard test. Wandering behaviour, as well as both risk-assessment tools, were able to identify fallers (Figure 1) and recurrent fallers (see Appendix 1 in the supplementary data on the journal’s website http://www.ageing.oxfordjournals.org) from the time of admission and throughout the first 21 days of patient stay.

Discussion

In this study and in this setting, the Downton and STRATIFY tools tested did not have as high a total accuracy as predictors of falls as the single observation of wandering behaviour. This confirms findings by other authors regarding the variable performance of established fall risk scores such as the STRATIFY. The performance of the tool in this study was inferior to that reported by the original tool authors. This is a repeated finding across various studies [6, 12], and at our current state of knowledge the search for the perfect fall risk-assessment tool remains elusive.

While the clinical observation of wandering gave the highest total predictive accuracy due to the significantly higher specificity, the sensitivity of this observation was low implying that a higher proportion of fallers were not identified as being at risk. As the majority of falls in hospital occur around the bed area a small number of patients with restlessness or agitation in bed but unable to wander might have been missed, possibly contributing to the lower sensitivity. Although the definition of wandering was based on standardised criteria, another limitation of the study that could have contributed to this low sensitivity was that the judgement about wandering was made by a clinician (medic) who had already used two risk-assessment tools, and this could have influenced the outcome of the wandering assessment. In addition, the assessment was reliant on the observations of nursing staff during out-of-hours.

It would not be advisable to substitute current falls risk assessment with a single observation of wandering as a single item assessment for falls risk assessment. In spite of the apparent higher accuracy, the fact that sensitivity was low means that many patients who eventually fell would not have been identified as being at risk. Current tools have low accuracy due to low specificities. Arguably, this is a good thing. In fact, a mark of good clinical practice is when patients at risk of falls are prevented from falling. The act of preventing a fall in such patients lowers the specificity, thus affecting the predictive accuracy of the tool. Trying to develop further tools that are expected to continue to deliver a high total predictive accuracy with high sensitivities and specificities in areas outside the place of validation is therefore a futile and pointless process. Perfection of prediction suggests that the performance of the tools and the purpose of their use are uniform across clinical settings. This is not so, and there is considerable evidence to suggest that this is the case [6–12].

It may well be time for a new appraisal of how risk-assessment tools are used and evaluated. Filling in tools that are not very accurate may not be an effective use of time. False reassurance or poor targeting of interventions across a whole load of people unlikely to fall are the inevitable result of
poor specificity or low positive predictive value, respectively. Considerable work has been done in identifying the common risk factors for falls in a hospital environment. These risk factors have consistently been shown to be associated with falls in that environment. In units that have high fall rates it might be more efficacious and productive to identify these common risk factors in all patients and implement individual care plans for all patients who have these risk factors.

**Key points**
- Clinical observation of wandering behaviour has a higher total accuracy than STRATIFY and Downton scores when used as a fall risk-assessment tool.
- Sensitivity of clinical observation alone is low implying that a higher proportion of fallers are not identified as being at risk. It is therefore not recommended that using the clinical observation of wandering should replace the current risk-assessment tools despite their limitations.
- As specificity of a tool changes with successful falls prevention it is futile trying to develop further tools that are expected to deliver a high specificity in areas outside their place of validation.

**Conflicts of interest**

None

**Supplementary data**

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

**References**

Fall risk-assessment tools in a rehabilitation ward


Received 4 September 2007; accepted in revised form 9 January 2008