Predicting the risk of functional decline in older patients admitted to the hospital: a comparison of three screening instruments

SIR—Almost 30% of older patients develop functional decline, while being hospitalised [1, 2]. This often irreversibly diminishes quality of life, results in an increased risk of subsequent illness or death and is associated with prolonged hospital stay, higher rates of institutionalisation, readmission and larger health care expenditure [3–5]. Predictors of functional decline are age, admission diagnosis, lower functional status, impaired cognitive status, comorbidities, polypharmacy and length of hospital stay [6–9]. Various screening tools, encompassing (some of) these predictors are available: the Triage Risk Screening Tool (TRST) [10, 11], the Identification of Seniors at Risk (ISAR) [12, 13] or the Variable Indicative of Placement risk (VIP) [14]. For an overview of their reported diagnostic characteristics, see Appendix 1 in the supplementary data available at Age and Ageing online.

The diagnostic characteristics of the ISAR and the TRST are primarily tested at the emergency department (ED), in a population of ambulant patients. The VIP, used on both the ED and other wards, is previously tested for predicting discharge problems and an increased length of stay but is not yet evaluated using functional decline as the primary outcome. However, the intended use in clinical practice is overlapping to such a degree that comparing their predictive accuracy for assessing functional decline in hospitalised older patients seems favourable.

Methods

Study population

In this longitudinal study, executed at a 1470-bed academic hospital, patients aged 65 years and above and hospitalised following admission to the ED were evaluated for eligibility. Patients were included if they were Dutch-speaking, and could be reached by telephone. Patients with significant cognitive decline or those unable to give oral informed consent (1.9%) were included when a primary caregiver was present providing proxy consent. From October 2005 to December 2005, 431 persons were asked to participate. A total of 117 patients were excluded; the remaining 314 were subsequently assessed. Out of these patients, 213 were hospitalised following their admission to the ED, and thus included in the study. Results in patients who were discharged after ED visit were reported elsewhere [15, 16]. For the exclusion details and a flow chart of patient selection, see Appendix 2 in the supplementary data available at Age and Ageing online.

Screening instruments

The TRST screens older persons for their risk of hospitalisation, functional decline, nursing home admission or readmission to the ED [10]. We used a modified Flemish version of the instrument comprising five yes/no topics: cognitive impairment; difficulty in walking, transferring; living alone with no available caregiver; polypharmacy; and recent hospitalisation [17]. The presence of cognitive impairment or two or more risk factors designates the older person as ‘high risk’ [10, 11].

The ISAR identifies older persons at risk for mortality, functional decline, readmission and institutionalisation [12, 13]. It comprises six self-report questions on functional dependence, recent hospitalisation, impaired memory and vision, and polypharmacy. Response to these items is dichotomous (e.g. yes/no). Patients with a score of two or more are considered to be at risk.

The VIP assesses older patients’ degree of independence, with the aim of detecting patients with potential discharge problems [14]. With three yes/no questions, the patients’ living situation, functional status and cognitive status are evaluated. A patient is considered at risk when at least two positive answers are reported.

For a detailed overview of the instruments’ items, see Appendix 3 in the supplementary data available at Age and Ageing online.

Functional decline

We measured functional decline using the Katz Index of Activities of Daily Living (ADL) [18]. An item is scored 1 if the patient is able to perform the activity without help and scored 0 if the patient needs help. The total score ranged from 0 (total dependence) to 6 (total independence) [14]. Functional decline was defined as a loss of at least one point on the ADL scale when compared to the premorbid ADL score [2, 16].

Procedure

Four trained data nurses who collected demographic and clinical variables administered the screening instruments and evaluated the patients’ premorbid functional status (i.e. 2 weeks before admission). All patients were phoned at 14, 30 and 90 days post-discharge to assess their current functional status. The study was approved by the Institutional Review Board.
The mean premorbid ADL score of participating patients was 5.0 (SD = 1.5). On admission, 14 and 30 days post-discharge the mean ADL was 3.0 (SD = 1.8), 4.5 (SD = 2.0) and 4.7 (SD = 1.8), respectively, which was always lower than the initial score \((P < 0.001)\). Ninety days post-discharge, the mean ADL (4.9; SD = 1.7) still differed from the premorbid score \((P = 0.01)\). Two weeks before admission to the ED, 61.0\% of patients (130/213) reported to be independent for all items on the Katz scale. On admission, 19.9\% of patients (42/211) were functioning independently, and 14, 30 and 90 days post-discharge 51.6\% (96/186), 56.8\% (104/183) and 58.9\% (99/168) of patients were independent, respectively.

On admission and at 14, 30 and 90 days post-discharge, 69.7\% (147/211), 28.5\% (53/186), 23\% (42/183) and 22.6\% (38/168) of patients experienced functional decline, respectively.

### Predictive accuracy of ISAR, TRST and VIP (Table 2)

The ISAR showed good sensitivity \((\geq 74\%)\) and a high NPV on all measurement points \((\geq 83\%)\). Specificity, PPV and accuracy were low \((\leq 36\%, \leq 33\%\) and \(\leq 49\%, \) respectively). Overall, the sensitivity of the TRST remained stable \((\geq 77\%)\) and its NPV was high \((\geq 84\%)\). Specificity, PPV and accuracy were low \((\leq 50\%, \leq 38\%\) and \(\leq 57\%, \) respectively). The VIP had a low sensitivity \((\leq 43\%)\) and a high specificity \((\geq 86\%)\). Its NPV was high \((\geq 76\%)\) and its PPV was moderate \((\geq 47\%)\). The overall accuracy, however, was good \((\geq 70\%)\).

ROC curve analyses revealed that the area under the curve \((AUC)\) was moderate for all three instruments, on all measurement points \((\geq 55\%\) and \(\leq 65\%).

### Discussion

The ISAR and the Flemish TRST are comparable regarding their diagnostic characteristics, with a slight benefit for the TRST. With an acceptable sensitivity and NPV, they appear to be good screening instruments for predicting functional decline after hospitalisation. Their specificity, however, is low, resulting in at least 50\% false positive scores. Using the TRST or ISAR to predict decline in older, hospitalised patients may thus lead to the implementation of extensive interventions in a considerable number of non-risk patients. The specificity and accuracy of the VIP were high, but its sensitivity was unacceptably low, which would result in missing up to 68\% of high-risk patients.

In previous studies, the TRST had a balanced sensitivity and specificity \([10, 11]\). Sensitivity rates in our study were higher, and specificity was up to 13\% lower. Although we used a modified Flemish version of the TRST \(\text{(i.e. changing the item ‘recent falls’ into the more specific ‘fall in past 6 months and omitting the ‘professional recommendations’ item because of its perceived subjectivity), the reason for these differences may lie in the specific characteristics of hospitalised patients (versus ambulatory patients), the different follow-up times and differences in sample size.}

For the ISAR, sensitivity and specificity was remarkably lower than reported earlier \([13]\). The above-mentioned reasons and another way of measuring functional decline
Table 2. Diagnostic values for the different screening instruments 14, 30 and 90 days after discharge

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Follow-up</th>
<th>Cut-off score</th>
<th>Area under ROC</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV</th>
<th>NPV</th>
<th>LR-P</th>
<th>LR-N</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISAR</td>
<td>14</td>
<td>≥ 2</td>
<td>0.584</td>
<td>0.81 (0.69–0.90)</td>
<td>0.36 (0.32–0.40)</td>
<td>33</td>
<td>83</td>
<td>1.26</td>
<td>0.53</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>≥ 2</td>
<td>0.570</td>
<td>0.79 (0.65–0.89)</td>
<td>0.35 (0.31–0.39)</td>
<td>27</td>
<td>85</td>
<td>1.22</td>
<td>0.60</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>≥ 2</td>
<td>0.549</td>
<td>0.74 (0.59–0.85)</td>
<td>0.36 (0.32–0.40)</td>
<td>25</td>
<td>83</td>
<td>1.15</td>
<td>0.71</td>
<td>45</td>
</tr>
<tr>
<td>TRST</td>
<td>14</td>
<td>≥ 2</td>
<td>0.627</td>
<td>0.77 (0.65–0.87)</td>
<td>0.48 (0.43–0.52)</td>
<td>38</td>
<td>84</td>
<td>1.45</td>
<td>0.47</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>≥ 2</td>
<td>0.626</td>
<td>0.78 (0.64–0.89)</td>
<td>0.47 (0.43–0.50)</td>
<td>30</td>
<td>88</td>
<td>1.48</td>
<td>0.47</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>≥ 2</td>
<td>0.640</td>
<td>0.78 (0.63–0.89)</td>
<td>0.50 (0.46–0.53)</td>
<td>31</td>
<td>89</td>
<td>1.57</td>
<td>0.43</td>
<td>56</td>
</tr>
<tr>
<td>VIP</td>
<td>14</td>
<td>≥ 2</td>
<td>0.588</td>
<td>0.32 (0.22–0.42)</td>
<td>0.86 (0.82–0.90)</td>
<td>47</td>
<td>76</td>
<td>2.23</td>
<td>0.79</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>≥ 2</td>
<td>0.654</td>
<td>0.43 (0.30–0.55)</td>
<td>0.88 (0.84–0.91)</td>
<td>51</td>
<td>84</td>
<td>3.53</td>
<td>0.65</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>≥ 2</td>
<td>0.622</td>
<td>0.37 (0.24–0.50)</td>
<td>0.88 (0.84–0.91)</td>
<td>47</td>
<td>83</td>
<td>3.10</td>
<td>0.72</td>
<td>76</td>
</tr>
</tbody>
</table>

ROC, receiver operating characteristic; PPV, positive predictive value; NPV, negative predictive value; LR-P, likelihood ratio of a positive test; LR-N, likelihood ratio of a negative test.

(Dendukuri et al. operationalised it as ‘severe functional impairment’, equalling dependency on three or more ADLs) could be the cause of this.

Although the VIP originally had a good accuracy for predicting discharge problems and increased length of stay [14], this could not be repeated with functional decline as the primary outcome.

A possible study limitation is the measurement of functional decline. We used a very strict definition in accordance with previous studies [3, 9]. A higher threshold for functional decline might have generated different results. The remaining study limitations can be found in Appendix 4 in the supplementary data available at Age and Ageing online.

Health care professionals may still benefit from integrating the TRST or ISAR in daily practice. It increases awareness regarding the basic geriatric attention points reflected by these instruments, and a positive screening result could be the motive for a targeted intervention by a geriatric consultation team [19]. ‘False positives’ could be filtered out based on the clinical expert-opinion of the consultation team’s nurse, optimising the quantitative screening result and avoiding the unfolding of an extensive and costly multidisciplinary intervention.

Further research should focus on derivation and validation of screening tools to correctly identify older people who require additional support after hospitalisation.

Key points

- Hospitalised, older patients are at risk for developing functional decline. Specific interventions could be undertaken, when high-risk patients are identified.
- This study compared the diagnostic characteristics of the ISAR, TRST and VIP for predicting functional decline in older, hospitalised patients.
- The study found that the ISAR and TRST seem to be good screening instruments, but their low positive predictive and high false positive values may result in poorly targeted interventions.

Acknowledgements

We would like to thank all patients who participated in this study and are grateful for the help of the nurse researchers Koen De Riddler and Katrien Geyskens.

Conflicts of interest

There are no conflicts of interest.

Supplementary data

Supplementary data are available at Age and Ageing online.

References


SIR—Inappropriate prescribing (IP) encompasses the use of medicines where the risk of an adverse drug event (ADE) outweighs the clinical benefit, particularly when safer or more effective alternatives are available [1, 2]. IP also includes the use of medicines that increase the likelihood of drug–drug and drug–disease interactions, the mis-prescribing of medicines (incorrect dose, frequency and duration) and the under-use of clinically indicated medicines [3–5]. IP is highly prevalent in older people and has been associated with preventable ADEs, hospitalisation, institutionalisation, death and resource wastage [6–12]. With increasing proportions of older people worldwide, quality and safety of prescribing are becoming a global healthcare concern [5, 13].

One way of identifying IP is to use prescribing indicators such as the recently validated STOPP (Screening Tool of Older Persons’ Prescriptions) and START (Screening Tool to Alert doctors to Right Treatment) criteria [14]. STOPP comprises 65 indicators for potentially inappropriate prescribing including drug–drug and drug–disease interactions, therapeutic duplication and drugs that increase the risks of cognitive decline and falls (Appendix 1 in the supplementary data at Age and Ageing online) [14]. START incorporates 22 evidence-based indicators for prescribing omissions in older people (Appendix 2 in the supplementary data at Age and Ageing online) [14]. STOPP/START criteria are organised according to physiological systems for ease of use. Their content validity was established by a Delphi consensus process in which 18 experts in geriatric pharmacotherapy from Ireland and the United Kingdom participated [14]. A recent study showed that 35% of 715 acutely ill older patients requiring hospitalisation were regularly prescribed at least one potentially inappropriate medication according to STOPP criteria and 12% of admissions were directly attributable to associated serious ADEs [15]. Another study of 600 older patients showed that 58% were not prescribed clinically indicated medications without contraindication according to START criteria [16].

Prospective randomised controlled trials are needed to test whether routine clinical application of STOPP/START criteria can significantly improve prescribing appropriateness and reduce drug-related morbidity. However, before demonstrating effects on patient outcome, a screening tool must be generalisable and reliable. Inter-rater reliability of STOPP/START criteria was substantial when tested between two researchers (kappa coefficient 0.75 STOPP criteria and 0.68 START criteria) [14]. Further evaluation of reliability between health professionals practicing in different countries is warranted to determine if STOPP/START criteria are generalisable. Accordingly, the aim of this study was to