Predictive validity of the START for unauthorised leave and substance abuse in a secure mental health setting: a pseudo-prospective cohort study

Laura E. O’Shea Geoffrey L. Dickens

This is the accepted manuscript © 2015, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International http://creativecommons.org/licenses/by-nc-nd/4.0/

The published article is available from doi:
http://dx.doi.org/10.1016/j.ijnurstu.2015.02.007
Predictive validity of the START for unauthorised leave and substance abuse in a secure mental health setting: a pseudo-prospective cohort study

Laura E. O'Shea\textsuperscript{ab}

Geoffrey L. Dickens\textsuperscript{a,c}

\textsuperscript{a} St Andrew’s Academic Department, St Andrew’s Hospital, Billing Road, Northampton, NN1 5DG. United Kingdom. Email: leoshea@standrew.co.uk

\textsuperscript{b} King's College London Institute of Psychiatry, Psychology & Neuroscience, 16 De Crespigny Park, London, SE5 8AF

\textsuperscript{c} School of Social and Health Sciences, Abertay University, Dundee, United Kingdom. Email: g.dickens@abertay.ac.uk

**Correspondence:** Geoffrey L. Dickens, School of Social and Health Sciences, Abertay University, Kydd Building, Bell Street, Dundee, DD1 1HG, UK. Email: g.dickens@abertay.ac.uk. Tel: 01382 308257.
What is already known about this topic?
- The START aims to assist practitioners with risk assessment for multiple adverse outcomes including unauthorised leave and substance abuse.
- Its predictive validity for aggression is well-established but it is less clear for relatively infrequent outcomes like unauthorised leave and substance abuse.

What this paper adds?
- START specific risk estimates for unauthorised leave and substance abuse are significant predictors of their respective outcomes.
- The high Negative Predictive Value of the tool suggests it is most useful for screening out low risk patients from more restrictive risk management interventions.
- Researchers should examine risk formulation for these outcomes among clinical teams because their structured judgement outperforms the quantitative element of the START,
Predictive validity of the START for unauthorised leave and substance abuse in a secure mental health setting: a pseudo-prospective cohort study

ABSTRACT

Background: Risk assessment and management is central to the nursing role in forensic mental health settings. The Short Term Assessment of Risk and Treatability (START) aims to support assessment through identification of risk and protective factors. It has demonstrated predictive validity for aggression; it also aims to aid risk assessment for unauthorised leave and substance abuse where its performance is relatively untested.

Objectives: To test the predictive validity of the START for unauthorised leave and substance abuse.


Settings: Four centres of a large UK provider of secure inpatient mental health services.

Participants: Inpatients resident between May 2011 and October 2013 who remained in the service for 3-months following assessment with the START by their clinical team. Exclusion criteria were missing assessment data in excess of prorating guidelines. Of 900 eligible patients 73 were excluded leaving a final sample size of N=827 (response rate 91.9%). Mean age was 38.5 years (SD=16.7); most participants (72.2%) were male; common diagnoses were schizophrenia-type disorders, personality disorders, organic disorders, developmental disorders and intellectual disability.

Methods: Routinely conducted START assessments were gathered. Subsequent incidents of substance abuse and unauthorised leave were coded independently. Positive and Negative Predictive Values of low and elevated risk were calculated. Receiver Operating Characteristic analysis was conducted to ascertain the predictive accuracy of the assessments based on their sensitivity and specificity.

Results: Patient-based rates of unauthorised leave (2.4%) and substance abuse (1.6%) were low. The Positive and Negative Predictive Values for unauthorised leave were 5.9% and 98.4%; and for substance abuse 8.1% and 99.0%. The START specific risk estimate for unauthorised leave predicted its associated outcome (Area under the curve=.659, p<.05, 95% CI .531, .786); the substance abuse risk estimate predicted its outcome with a large effect size (Area under the curve=.723, p<.01, 95% CI .568, .879).

Conclusions: The study provides limited support for the START by demonstrating the predictive validity of its specific risk estimates for substance abuse and unauthorised leave. High Negative Predictive Values suggest the tool may be of most utility in screening out low risk individuals from unnecessary restrictive interventions; very low Positive Predictive Values suggest caution before implementing restrictive interventions in those rated at elevated risk. Researchers should investigate how multidisciplinary teams formulate risk assessments for these outcomes since they outperform the quantitative element of this tool.

Keywords: psychiatric nursing, risk assessment, risk management, sensitivity and specificity, substance abuse
INTRODUCTION

The assessment and management of risk in secure and forensic mental health settings is central to the role of mental health and learning disability nurses (Bowen & Mason, 2012; Doyle & Dolan, 2008; Fluttert, Van Meijel, Webster, Nijman, Bartels, & Grypdonck, 2008; Kettles & Woods, 2006; Mason & Phipps, 2010; Rask & Aberg, 2002; Timmons, 2010). Risk assessment has traditionally focused on the prediction of aggression and violence, but outcomes like unauthorised leave and substance abuse have significant potential to interrupt patient’s recovery and jeopardise service delivery (Bowers, Douzenis, Galeazzi, Forghieri, Tsopelas, Simpson, & Allan, 2005; Isaac, Isaac, & Holloway, 2005; Muir-Cochrane, & Mosel, 2008; Scott, Whyte, Burnett, Hawley, & Maden, 2004). Accurate prediction of those most likely to engage in these adverse outcomes, therefore, holds the potential to inform targeting of resources and interventions to ameliorate risk.

Tools that aim to assist with the assessment of risk in forensic and secure settings have generally centred upon the identification of empirically-derived, static and dynamic risk factors. More recently, there has been a growing recognition of the need to consider the protective factors that individual’s hold; these have been defined as “conditions or attributes of individuals, families, communities, or the larger society that reduce or eliminate risk” (Child Welfare Information Gateway et al, 2014). The consideration of protective factors may have clinical advantages in terms of informing a more rounded perspective of patients, and promoting the therapeutic relationship (de Vogel et al, 2012); informing risk management and treatment plans (Nonstad et al, 2010); and reducing bias and consequent overestimation of risk that may arise from concentration solely on risks, or vulnerabilities, rather than considering an individual’s personal strengths (de Ruiter & Nicholls, 2011). At the same time, there has been a suggestion
that concentrating only on risk outcomes related to violence and aggression is inappropriate when, in fact, the care and management of patients in inpatient settings requires the consideration of a much broader range of clinical issues (Webster, Martin, Brink, Nicholls, & Desmarais, 2009).

The Short Term Assessment of Risk and Treatability (START; Webster et al, 2009) was developed in part to redress the criticisms about risk assessment outlined above. The START requires consideration of a range of factors both in terms of the patients’ risks (termed ‘Vulnerabilities’) and their protective factors (termed ‘Strengths’). The underlying assumption of the model is that people can hold both risk and protective factors in key areas concurrently. In contrast, other models variously view a protective factor simply as the absence of a risk factor (Costa, Jessor & Turbin, 1999), or as a factor that lies at the opposite end of a continuum to a risk factor (Hawkins, Catalano & Miller, 1992). As a result, raters of the START are required to make separate judgments about both strengths and vulnerabilities. Further, as with other structured professional judgment tools, assessors are instructed to make an overall specific risk estimate for each outcome of interest representing the raters’ view of the overall level of risk; these are rated as low, moderate or high. The START aims to facilitate clinicians in assessment of risks for a range of adverse outcomes in addition to aggression and violence: namely, self-harm, suicidality, self-neglect, victimization, unauthorised leave and substance abuse. Meta-analysis of the existing research literature on the START has revealed that the tool’s Strength scale, Vulnerability scale, and specific risk estimate for violence have good validity for the prediction of aggressive outcomes (O’Shea & Dickens, 2014). Meta-analysis also suggests that the specific risk estimate for self-harm, but not the Strength or Vulnerability scale totals, predicts its associated outcome. Predictive ability for other adverse outcomes is less well supported. More
specifically, the ability of the tool’s Strength and Vulnerability scales to successfully predict unauthorised leave and substance abuse has been reported in only one study (Braithwaite, Charette, Crocker, & Reyes, 2010) prior to O’Shea and Dickens’ (2014) meta-analysis. Since then, O’Shea, Picchioni, and Dickens (2014) aimed to investigate the predictive validity of the START for the full range of adverse outcomes. Results for aggression and self-harm mirrored those of the meta-analysis; the specific risk estimate for victimisation predicted its corresponding outcome, whilst the Strength scale significantly predicted self-neglect. However, reflecting the secure study setting, base rates for unauthorised leave (3%) were low but in accordance with the literature which suggests rates of 1-4% (Brook, Dolan, & Coorey, 1999; Dolan & Snowden, 1994; Huws & Shubsachs, 1993; Moore, 2000). The rate of active substance abuse (1.5%) was very low, and considerably lower than in Braithwaite et al’s (2010) study which reported on predictive ability of the START in a civil psychiatric hospital setting. Therefore, our previous study was under-powered to detect statistically significant predictive ability of the START for these outcomes even though the sample was the largest, to date, in the research literature about the tool.

**Aims of the current study**

This study aimed to extend our previous study (O’Shea et al, 2014), which examined the predictive validity of the START for aggression, self-harm/suicide, victimisation and self-neglect, by investigating the predictive efficacy of the START for the remaining outcomes it intends to predict, namely unauthorised leave and substance abuse. Demonstrating positive predictive ability of tool for these outcomes in an adequately powered sample would provide an evidence-base for the use of the START for these additional outcomes that is currently lacking.

**METHOD**
Participants

St Andrew’s is a charitable provider of secure mental health inpatient care located at four sites in England. Eligible participants were all patients resident in the services on 1st May 2011, and all of those subsequently admitted until 31st October 2013, who had at least one START risk assessment completed and who then remained in the service for at least the next three months. A subsample of these patients admitted between May 2011 and July 2012 (n=200) were previously reported on by O’Shea, Picchioni & Dickens (2014). Patients were excluded if their START assessment had in excess of five missing Strength or five missing Vulnerability ratings as per the prorating guidelines in the START manual (Webster et al, 2009). We used MedCalc for Windows, version 13.3.3 (MedCalc Software, Ostend, Belgium) to calculate the sample sizes required to detect area under the curve values commonly reported as a large effect size (.75; Dolan & Doyle, 2000) for both unauthorised leave and substance abuse based on rates detected in our previous study (3% and 1.5% respectively). Setting α=0.05 and β (1-Power) = 0.20, the required sample sizes (positive cases required/negative cases required) was 11/356 for unauthorised leave and 11/723 for substance abuse.

Procedure

NHS research ethics approval was not required for the study since all data had identifying details redacted prior to use by the researchers. We used a pseudo-prospective cohort design. START assessments were completed by the patients’ multidisciplinary team during admission, as part of routine clinical practice. Additionally, risk incidents were recorded by clinical staff in electronic progress notes on each shift. Each patients’ first START assessment, demographic and clinical data, and risk incidents for the three months following START assessment were extracted from their records, anonymised, and linked by a unique code number.
This study was reported following the Risk Assessment Guidelines for the Evaluation of Efficacy (Singh, Yang, Mulvey, & The RAGEE Group, 2014), which are designed to facilitate reporting of studies investigating the predictive efficacy of violence risk assessments. The majority (38; 76%) of these guidelines were met. There were eight items that were not applicable; four items relating to subgroup/post hoc analyses, three items regarding disclosures of commercial interest, sources of funding, authorship and previous publications, and one item relating to tool modification. Due to the pseudo-prospective nature of this study and reliance on routinely collected data, three items were not met and one was partially met; it was not possible to report the index offences of included patients, identify the sources of information used to administer the instrument, state the number of assessors involved in completing START assessments, or report inter-rater reliability of the risk assessments.

Measures

START Assessment. The START is a structured professional judgment tool that aims to assist clinicians in assessing the likelihood of a range of adverse outcomes occurring in patients with mental and personality disorders. It comprises 20 dynamic items scored by a multidisciplinary team on two 3-point scales, one for risk factors (Vulnerability scale) and one for protective factors (Strength scale): 0 indicates no/minimal vulnerability or strength evident, 1 indicates moderate vulnerability/strength, and 2 indicates high vulnerability/strength. There is no suggested cut-off score on either scale indicating risk level; instead, raters are instructed to use scores alongside idiographic factors to make specific risk estimates (low, moderate, or high) about the likelihood of each of seven risk outcomes occurring within a maximum period of three months. The two specific risk estimates of interest in the current study were for unauthorised leave and substance abuse. Total scores on the Strength and Vulnerability scales were prorated to
account for missing items in accordance with guidelines in the START manual (Webster et al., 2009). The START is intended to be completed by a team of mental health specialists in order to incorporate a range of professional opinions. There are no specific recommendations about the number or precise qualifications of multidisciplinary raters, or their required level of training in the START assessment. In the current study setting, a 1-day structured training session in START assessment is provided to all clinicians covering theoretical and practical aspects of the tool. Training includes team discussion and rating of pseudonymised clinical cases; reliability is enhanced through the provision of ratings given by START experts and teams from previous training sessions. Completed START assessments are then signed off by three members of the multidisciplinary clinical team from different professions (nurse, psychiatrist, psychologist, occupational therapist, social work). Each patients’ START assessment is reviewed every 3 months and a sample of cases is regularly audited to ensure compliance (see Appendix for START coding sheet).

Demographic and Clinical Data. Information was extracted regarding patients’ age, gender, admission/discharge date, level of security, legal status, self-reported ethnicity and International Statistical Classification of Diseases and Related Health Problems 10th Revision (World Health Organisation, 1992) psychiatric diagnoses, as recorded by the patients’ consultant psychiatrist.

Risk Outcomes. Hospital policy dictates that an electronic progress note is entered for each patient by a qualified member of clinical staff on a per shift basis. Following the entry of each progress note, staff electronically flag the note if any of a range of risk outcomes has occurred. We collated all incidents flagged as ‘absconding’ or ‘substance misuse’ (the local terms used for the two relevant outcomes). Progress notes for entries flagged as absconding were
Predictive validity of the START

coded using the START Outcome Scale (SOS; Nicholls et al., 2007). The SOS comprises 12 outcome categories, rated on a criterion-referenced severity scale of 0 (outcome absent) to 4 (most severe) including unauthorised leave and substance abuse. For unauthorised leave, the raters were required to judge whether each flagged note met the requirement for a level 1 incident or above. We did not use the SOS to code substance abuse since the narrative anchors comprise descriptions of global outcome (e.g., “Frequent substance use leading to significant impairment”) rather than of single incidents of use. For substance use, therefore, we simply coded flagged entries as either ‘substance abuse present’ or ‘substance abuse absent’. We also noted whether the substance involved was alcohol or another illicit substance. We included incidents of confirmed substance use (e.g., positive drugs screening test or breathalyser), and incidents where staff expressed suspicion of current substance abuse. We did not include incidents where staff reported discussions with patients about historical substance abuse or current substance-related ideation. Rating was conducted by both authors, who were blind to the results of the START assessments at the time of coding, and inter-rater reliability was established by on a sample of flagged entries for 20 patients for each outcome.

Data Analysis

Inter-rater reliability for coding of risk outcomes was tested using the kappa (k) statistic. Descriptive statistics were used to describe sample characteristics, the distribution of START scores and specific risk estimates, and the occurrence of risk outcomes. Differences in mean scores between those who had and had not engaged in each outcome were examined through use of independent t tests. Associations between the Strength scores, Vulnerability scale scores and specific risk estimates were examined using Spearman’s ρ correlations; coefficients of .10, .30, and .50 are considered small, moderate, and large, respectively (Cohen, 1992). Positive
predictive value and negative predictive value of the specific risk estimates for unauthorised leave and substance abuse was calculated to provide an index of the tool’s performance in terms of identification of true positive and true negative predictions; this can assist in determining whether a tool is of greater value for screening out those individuals at low risk from further assessment or whether its value lies in identifying higher risk individuals. We also calculated the number needed to detain, which is the number of high risk individuals that would need to be detained in order to prevent a single incident occurring in the community, and the number safely discharged, which is the number of low risk individuals that could be discharged prior to a single incident occurring in the community (DeClue & Campbell, 2013; Fazel, Singh, Doll & Grann, 2012). However, to reflect the inpatient setting, number needed to detain should be considered as the number of high risk individuals who require risk management strategies targeting unauthorised leave or substance abuse to prevent one incident occurring within the inpatient setting; number safely discharged corresponds to the number of low risk individuals that could safely be managed without strategies targeting these outcomes prior to an incident occurring. To do this we assigned those rated at elevated risk (moderate or high) as test outcome positive, and those rated at low risk as test item negative. Predictive validity was investigated by calculating area under the curve values using Receiver Operating Characteristic analysis; receiver operating characteristic curves are a plot of sensitivity (proportion of true positives that are correctly identified; Kirkwood & Sterne, 2003) against 1-specificity (proportion of true negatives that are correctly identified; Kirkwood & Sterne, 2003). The total Strength score was inverted before conducting receiver operating characteristic analysis to facilitate comparisons with the predictive efficacy of the Vulnerability scale scores and specific risk estimates. Area under the curve values range from 0 to 1 with .5 representing a level of chance prediction. Area under the curve values
of .556, .638 and .714 correspond to small (.2), moderate (.5) and large (.8) Cohen’s d values respectively (Cohen, 1992), which are one of the most commonly reported measures of effect size (Kraemer & Kupfer, 2006; Rice & Harris, 2005). Odds ratios were also calculated to present the increase in odds for each one point increase on the Strength and Vulnerability scales, and between those categorised as moderate or high risk, compared to low risk, by the specific risk estimates, for each adverse outcome occurring. All analyses were conducted using SPSS version 18 for Windows (SPSS, 2009).

RESULTS

Inter-rater reliability of progress note coding

Inter-rater reliability for coding the outcomes from progress notes was in the excellent range; substance abuse k=.894; unauthorised leave k=.857.

Participants

Of the 900 patients that met the inclusion criteria, 73 were excluded due to excessive missing data from the START assessment leaving a final sample of N=827 (response rate 91.9%) with a mean age of 38.5 years (SD =16.7). There were 597 (72.2%) males and 230 females (27.8%). About a third of the sample had no recorded ethnicity (n=255, 30.8%), 467 (56.5%) were Caucasian, 29 (3.5%) were of Asian or Asian British ethnicity, 43 (5.2%) were Black or Black British, and the remainder (n=33, 4.0%) were of mixed ethnic background. Patients were detained under forensic (n=379, 45.8%) and civil (n=381, 46.1%) sections of the Mental Health Act 1983 (amended 2007) while n=67 (8.1%) were informal. The majority of patients (n=547, 66.1%) were resident in low secure or locked wards and 267 (32.3%) were in medium secure wards; the remaining patients (n=13, 1.6%) were on an open ward. The mean time between
admission and START assessment was 892 days \((SD = 1709.4)\). The psychiatric diagnoses of the sample are presented in Table 1.

**Table 1:** Diagnostic characteristics of the sample

<table>
<thead>
<tr>
<th>International Statistical Classification of Diseases and Related Health Problems 10(^{th}) Revision Category</th>
<th>(n)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(F20–F29) schizophrenia, schizotypal and delusional disorders</td>
<td>332</td>
<td>40.1</td>
</tr>
<tr>
<td>(F60–F69) disorders of adult personality and behaviour</td>
<td>267</td>
<td>32.3</td>
</tr>
<tr>
<td>(F00–F09) organic, including symptomatic, mental disorders</td>
<td>215</td>
<td>26.0</td>
</tr>
<tr>
<td>(F80–F89) disorders of psychological development</td>
<td>151</td>
<td>18.3</td>
</tr>
<tr>
<td>(F70-F79) mental retardation</td>
<td>146</td>
<td>17.7</td>
</tr>
<tr>
<td>(F10–F19) mental and behavioural disorders due to psychoactive substance use</td>
<td>101</td>
<td>12.2</td>
</tr>
<tr>
<td>(F30–F39) mood [affective] disorders</td>
<td>56</td>
<td>6.8</td>
</tr>
<tr>
<td>(F90–F98) behavioural and emotional disorders with onset usually occurring in childhood and adolescence</td>
<td>50</td>
<td>6.0</td>
</tr>
<tr>
<td>(F40–F48) neurotic, stress-related and somatoform disorders</td>
<td>36</td>
<td>4.4</td>
</tr>
<tr>
<td>(F50–F59) behavioural syndromes associated with physiological disturbances and physical factors</td>
<td>12</td>
<td>1.5</td>
</tr>
</tbody>
</table>

*Note.* Percentages ≠100% due to comorbidities

**Risk outcomes**

There were 23 incidents of unauthorised leave in the follow-up period involving 20 (2.4%) individuals; most (19, 82.6%) were relatively serious episodes where the individual’s whereabouts was unknown for a period of time. There were 22 incidents of substance abuse in the follow-up period involving 13 (1.6%) individuals; four incidents engaged in by four
individuals involved use of alcohol while 18 incidents engaged in by nine individuals involved other illicit substances including cannabis, benzodiazepines, and amphetamines. Those who engaged in any unauthorised leave during follow-up were significantly younger than those who did not \( (M = 27.10 \text{ years } [SD=16.72] \text{ v. } 38.77 \text{ years } [SD= 10.70], r=4.738, p<.001) \); those who engaged in substance abuse were younger than those who did not but this result was not statistically significant. All but one of the patients who engaged in any substance abuse were male while seven (35%) of those who engaged in unauthorised leave were female; neither of these results were statistically significant. Incidents of substance abuse and unauthorised leave occurred in all levels of security. No informal patients engaged in either risk outcome and there was no greater likelihood for patients detained under civil or forensic sections of the Mental Health Act to do so.

**START scores and risk estimates**

Scores on the Strength scale ranged from 0 to 40 \( (M=16.40, SD=7.40) \) and the Vulnerability scale ranged from 2 to 40 \( (M=24.48, SD=7.51) \). The two scale scores were significantly negatively correlated \( (p=-.694, p<.001; 95\% \text{ Confidence Interval [CI]} -.652, -.736) \). The mean Vulnerability scale scores were significantly higher for patients who had engaged in any unauthorised leave than for those who had not; however, mean Strength scores did not differ between the two groups (see Table 2). There was no significant difference in either Vulnerability or Strength scores between those who had and had not engaged in any substance abuse during follow-up.

The distribution of the specific risk estimates is presented in Table 3. The most common classification was low risk for both outcomes: unauthorised leave had the largest proportion of high risk classifications (8.2%). The specific risk estimate for unauthorised leave was
significantly positively correlated with the total Vulnerability scale score ($\rho = .374, p < .01; 95\% CI .304, .444$) and significantly negatively correlated with the total Strength scale score ($\rho = -.277, p < .01; 95\% CI -.205, -.350$); similarly, the specific risk estimate for substance abuse correlated with Vulnerability scale total ($\rho = .264, p < .001; 95\% CI .191, .338$) and Strength scale total ($\rho = -.160, p < .01; 95\% CI -.082, -.239$). Specific risk estimates for unauthorised leave and substance abuse were also significantly correlated ($\rho = .300, p < .01; 95\% CI -.212, -.388$).

**Table 2:** Mean Strength and Vulnerability scale scores for those who had engaged in risk outcomes vs. those who had not engaged.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Mean Strength Score (SD)</th>
<th>Mean Vulnerability Score (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Any unauthorised leave</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>15.25 (7.22)</td>
<td>27.75 (6.21)</td>
</tr>
<tr>
<td>No</td>
<td>16.43 (7.41)</td>
<td>24.40 (7.52)</td>
</tr>
<tr>
<td>$t$</td>
<td>0.720</td>
<td>-2.37*</td>
</tr>
<tr>
<td><strong>Any substance abuse</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>16.69 (6.20)</td>
<td>25.69 (6.86)</td>
</tr>
<tr>
<td>No</td>
<td>16.39 (7.43)</td>
<td>24.46 (7.52)</td>
</tr>
<tr>
<td>$t$</td>
<td>-.172</td>
<td>-.641</td>
</tr>
</tbody>
</table>

*$p < .05$

**Table 3:** Risk levels assigned by the START specific risk estimates

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unauthorised leave</strong></td>
<td>428 (51.8%)</td>
<td>137 (16.6%)</td>
<td>68 (8.2%)</td>
<td>194 (23.4%)</td>
</tr>
<tr>
<td><strong>Substance abuse</strong></td>
<td>504 (60.9%)</td>
<td>73 (8.8%)</td>
<td>26 (3.1%)</td>
<td>224 (27.1%)</td>
</tr>
</tbody>
</table>

**Predictive validity of the START**
The positive predictive value of the specific risk estimate for unauthorised leave was 5.9% and the negative predictive value for the same outcome was 98.4%. The positive predictive value of the specific risk estimate for substance abuse was 8.1% and the negative predictive value 99.0%. Numbers needed to detain were 17 for unauthorised leave and 12 for substance abuse; numbers safely discharged were 60 and 100 for unauthorised leave and substance abuse, respectively. The area under the curve values derived from receiver operating characteristic analysis are presented in Table 4. The total Strength scale scores predicted neither of the two outcomes at above-chance levels; the total Vulnerability scale score significantly predicted unauthorised leave producing a moderate effect size compared against equivalent Cohen’s $d$ values, but did not predict substance abuse. The specific risk estimates for unauthorised leave and substance abuse both predicted their respective outcomes; the area under the curve effect sizes were moderate and large respectively. Item-outcome analysis revealed that no single START item was a significant predictor of unauthorised leave. The START items ‘Substance Use’ and ‘Self-care’ were both significant predictors of substance abuse during follow-up in both their Strength and Vulnerability iterations. Combination of the protective scores for these two items did not result in a composite item with significant predictive validity; similarly, combination of the risk scores for the two items did not produce a useful composite item. Odds ratios were largely consistent with the results of the receiver operating characteristic analysis (see Table 5). Increases in total Vulnerability and Strength scores were associated with very small, but significant, increases in the odds of both outcomes occurring. Those given elevated specific risk estimates for unauthorised leave (i.e., moderate or high) were both around four times more likely to engage in the associated risk outcome than individuals assigned to the low risk category. For unauthorised leave, those assigned high risk were no more likely to engage in
this behaviour than those rated as moderate risk. For substance abuse, those rated at moderate risk were more than ten times more likely to abuse substances than those rated as low risk. Only one individual rated as high risk for substance abuse engaged in this outcome; odds ratios for those rated high risk compared with those rated as moderate or low produced non-significant results.

Table 4: Predictive validity of the START scores and specific risk estimates

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Strength Score</th>
<th>AUC</th>
<th>p</th>
<th>95% CI</th>
<th>Vulnerability Score</th>
<th>AUC</th>
<th>p</th>
<th>95% CI</th>
<th>Specific Risk Estimate</th>
<th>AUC</th>
<th>p</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unauthorised Leave</td>
<td>.545</td>
<td>.494</td>
<td>.530</td>
<td>[.426, .663]</td>
<td>.639</td>
<td>&lt;.05</td>
<td>.659</td>
<td>[.530, .748]</td>
<td></td>
<td>.659</td>
<td>&lt;.05</td>
<td>[.531, .786]</td>
</tr>
<tr>
<td>Substance Abuse</td>
<td>.484</td>
<td>.847</td>
<td>.341</td>
<td>[.341, .627]</td>
<td>.547</td>
<td>.557</td>
<td>[.403, .692]</td>
<td>.723</td>
<td>.01</td>
<td>[.568, .879]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Odds ratios for START scores and specific risk estimates

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Vulnerability Score OR 95% CI</th>
<th>Strength Score OR 95% CI</th>
<th>Specific risk estimate Moderate-Low OR 95% CI</th>
<th>Specific risk estimate High-Low OR 95% CI</th>
<th>Specific risk estimate High- Moderate OR 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unauthorised Leave</td>
<td>1.10* [1.00, 1.20]</td>
<td>.96 NS [.88, 1.04]</td>
<td>3.73*** [1.33, 10.48]</td>
<td>3.76*** [1.07, 13.20]</td>
<td>.992 NS [.29, 3.42]</td>
</tr>
<tr>
<td>Substance Abuse</td>
<td>1.10* [1.01, 1.20]</td>
<td>.96 NS [.88, 1.04]</td>
<td>10.59*** [3.27, 34.31]</td>
<td>3.99 NS [.45, 35.47]</td>
<td>2.65 NS [.31, 22.66]</td>
</tr>
</tbody>
</table>

*p<.05, ***p<.001, NS=Not significant

DISCUSSION

The current study has further clarified that the Vulnerability scale of the START has some predictive validity for unauthorised leave of patients resident in secure mental health services. Neither scale predicted substance abuse. These findings suggest that the Vulnerability scale may be considered a multi-factorial risk factor for unauthorised leave but not for substance
abuse, and the Strength scale is not a risk factor for either outcome. Item-outcome analysis revealed that no single START item predicted unauthorised leave; this reinforces that the Vulnerability scale, while significantly predictive, does not represent a potent risk factor. Item-outcome analysis revealed that the START item ‘Substance use’ was the best predictor of substance abuse in follow-up in both its Strength and Vulnerability iterations. In addition, and importantly, the specific risk estimates for unauthorised leave and substance abuse were significantly predictive of their respective outcomes, the latter achieving a large effect size. For unauthorised leave, individuals rated at elevated risk (moderate or high) were both around four times more likely to engage in the behaviour than other patients. For substance abuse, those rated at moderate risk were ten times more likely to engage in the outcome; odds ratios for those rated high risk were non-significant, probably reflecting that only one patient rated at high risk actually went on to abuse substances.

The finding that specific risk estimates were significantly predictive of their related outcome offers some qualified support for the START since the tool requires clinical teams to make a judgement about overall risk based on consideration of Strength and Vulnerability items, risk history, additional case-specific risk factors, and any other relevant issues. In particular, high negative predictive values for both unauthorised leave and substance abuse suggests that the main value of the specific risk estimates for both outcomes in clinical practice will be to screen out those judged to be at low risk from further lengthy assessment and restrictions designed to minimize risk. Conversely, very low positive predictive values indicate that the specific risk estimates identify many false positive cases and therefore careful consideration and further assessment of those rated at elevated risk should be conducted prior to further restriction.

However, while the specific risk estimates may have clinical utility, the current observational
study design precludes conclusions about the extent to which any element of the START or its guidelines contributed towards them. Alternatively, clinical teams may simply make a specific risk estimate based on experience and/or guesswork. Given that specific risk estimates for both outcomes were significantly associated with Strength and Vulnerability total scores, we can speculate that consideration of individual START items makes some overall contribution to the specific risk estimates. This is supported by the fact that teams are instructed to rate items prior to making the specific risk estimate; thus, the alternative explanation, i.e., that overall estimation of risk informs item rating, is less likely. Correlations between specific risk estimates and Strength and Vulnerability total scores, while statistically significant, did not achieve large effect sizes; therefore, we should conclude that specific risk estimates are not simply a translation of total score to overall risk level. Presumably, clinical teams are considering other sources of evidence not contained in the START scales and this has implications for future research. Given that, in item-outcome analysis, very few START items were predictive of the outcomes under consideration, we can conclude that there could be considerable scope to improve the empirical basis of the START for these outcomes. The START manual details the thinking behind the inclusion of each of its items but rarely draws on literature beyond that supporting the link between the item and violent and aggressive outcomes. It is possible that supplementation of START coding advice for relevant items with elements from the emerging literature on dynamic risk factors for absconding (Bowers, Jarrett & Clark, 1998; Muir-Cochrane & Mosel, 2008) may increase the potency of those items for predicting unauthorised leave. For example, coding instructions for the START item ‘Coping’ might emphasise the relationship between absconding and the patient’s ability to cope with disappointment and bad news (Bowers, Jarrett, Clark, Kiyimba, & McFarlane, 2000). There is considerably less evidence about substance abuse during
admission to inform assessment, and that which does exist largely considers relatively static factors such as age, gender and ethnicity (Bowers & Jeffery, 2008). There is contradictory evidence about the role of rule adherence as a factor for substance abuse (Alterman, Erdlen, Laporte, & Erdlen 1982; Alterman, Erdlen, McLellan, & Mann 1980; Alterman, Erdlen, & Murphy 1981). There is also some evidence that substance abuse in inpatients is related to START outcomes including aggression (Alterman, Erdlen, Laporte, & Erdlen 1982; Alterman, Erdlen, McLellan, & Mann 1980; Alterman, Erdlen, & Murphy 1981), self-harm (Kamali, Kelly, Gerbvin, Browne, Larkin, & O’Callaghan 2000), and unauthorised leave (Bowers, Simpson, & Alexander 2003); as a result, further advances on development of items for these outcomes holds the potential to improve the predictive ability of the items for substance abuse. Both outcomes considered in the current study represent failures of security, particularly of procedural and relational elements. However, substance abuse can probably be assumed to be a largely invisible outcome in secure mental health care in that it may well go undetected; care staff may perceive that substance misuse occurs but may lack concrete evidence (e.g., Dolan & Kirwan, 2001). In contrast, unauthorised leave is a highly visible outcome and, when patients are unaccounted for, unlikely to go undetected or unreported. Nevertheless, our finding that specific risk estimates predicted substance abuse suggests that team ratings of risk hold some validity. The current study results suggest that even a rating of moderate risk of substance abuse should be taken seriously since those rated at moderate risk were considerably more likely to engage in substance abuse than those rated low risk. Item-outcome analysis linking substance abuse with the START item substance abuse should also sensitize clinical teams to the need to work with patients with a history of substance abuse problems in order to ameliorate risk. The relationship between the
START item self-care and substance abuse is difficult to fathom and may simply result from multiple testing.

**Outcome base rates**

Results of the current study contrast somewhat with findings by Braithwaite et al (2010), from a study conducted in a civil psychiatric setting, who reported much higher behaviour base rates (a patient-based rate of 38.2% over 30 days for both outcomes), and significantly predictive area under the curve values for both scales for both behaviours. Like our study, Braithwaite and colleagues reported a significantly predictive specific risk estimate for substance abuse. We suggest that the secure setting investigated in the current study is sufficiently different from that in Braithwaite’s study to render any comparison meaningless. Rates of unauthorised leave in the current study falls within the 1-4% of all admissions range reported in many studies of absconding from secure and forensic services (Brook et al., 1999; Dolan & Snowden, 1994; Huws & Shubsachs, 1993; Moore, 2000). While the rate in our study is less than a more recent report of absconding from a forensic mental health service (14.4% of at risk patients; Wilkie, Penney, Fernane, & Simpson, 2014) there was a 2-year maximum window in that study compared with the 90-days in the current study. It is more difficult to establish whether base-rates of substance abuse were representative of the reality of life in a secure hospital setting.

Staff in secure units have reported that patients use illicit drugs (Dolan & Kirwan, 2001); and review of case notes suggests that about half of patients in forensic care report ‘current substance use’ (Scott, Whyte, Burnett, Hawley & Maden, 2004), but it is not clear how many patients are actually using substances while in inpatient secure care. In general psychiatric inpatient care, use of illicit substances during the first two weeks of admission was 9.7% in the UK (Bowers et al., 2005), but a number of studies suggest that substance abuse is reduced by patients during
inpatient care (e.g., Isaac, Isaac, & Holloway 2005; Phillips & Johnson 2003); a finding likely to be replicated and amplified in secure care.

**Limitations**

The reliance on pre-existing, routinely collected information limited the current study; however, it allowed us to collect sufficient data for an adequately powered study to detect predictive validity of the START for relatively rare outcomes. We could not verify diagnoses ourselves through structured methods and we were missing a large amount of data pertaining to ethnicity due to this being a self-reported variable. We were also missing approximately a quarter of the specific risk estimates for both unauthorised leave and substance abuse. A further limitation is that the START assessments were completed by the patients’ multidisciplinary team rather than researchers. It is possible that different results may have been obtained if completed by members of the research team with established inter-rater reliability. However, as the START is intended for use by a clinical team, this would not have provided an accurate representation of its performance as part of clinical practice. The analyses employed in the current paper are unable to account for the successful prevention of risk outcomes and this may contribute to smaller area under the curve values than are found when researchers undertake the risk assessment. Further, positive predictive values may be an underestimation of the proportion of elevated risk individuals that would have engaged in the outcomes if outcomes had not been successfully managed. However, in secure settings it is likely that risk assessment and management procedures would be in place in any event irrespective of who completed the research risk assessment. The SOS lacks a strong evidence base, but there is a lack of tools in general for measuring non-aggressive outcomes in forensic settings. In the end, we chose not to
use the SOS outcome for substance abuse but devised our own dichotomous variable and were able to code this reliably.

Implications and future considerations

The START specific risk estimates predicted unauthorised leave and substance abuse, target outcomes that appear, uniquely, to be within the operational scope of this tool. While the precise contribution of the START assessment process cannot be determined or quantified this provides some evidence for its new or continued use in secure mental health services. Predictive validity of the Vulnerability scale for unauthorised leave and of the Substance abuse item for substance abuse further supports this. However, because of the small numbers of items related to the outcomes there are few implications in terms of treatment targets for these outcomes. Given that clinical teams are managing to predict risk for these outcomes over and above the START scales and items then one useful avenue for future research would be to explore, in depth, the decision-making process that teams undergo when assessing risk for unauthorised leave and substance abuse. Findings could be combined with specific unauthorised leave and substance abuse risk and protective factors to develop or refine the START items to better reflect all of the outcomes that it aims to address. Further research to establish the reliability and validity of new items would, of course, be merited. Our findings should highlight to clinical teams the need to very carefully address the risk management plan for everybody rated at elevated risk for these outcomes since the high risk rating seems to confer little extra predictive ability over that of moderate risk. Given the limited opportunity for unauthorised leave and substance abuse in a secure psychiatric setting, the current findings should not be generalised outside of secure psychiatric settings without replication. Finally, we acknowledge that the ultimate aim of risk assessment tools should be to facilitate individualised packages of care and treatment that lead to
the reduction of adverse outcomes; therefore any new developments should be rigorously tested as interventional procedures.
REFERENCES


Predictive validity of the START

the assessment of protective factors for violence risk (English version 2nd Edition).
Utrecht: Forum Educatief.


Psychiatry, 5, 275–286.


predict violence and antisocial behaviour in 73 samples involving 24 827 people:
systematic review and meta-analysis. British Medical Journal, 345,

Risk management by early recognition of warning signs in patients in forensic psychiatric

and other drug problems in adolescence and early adulthood: implications for substance


APPENDIX

START Summary Sheet (Webster et al., 2009): reproduced with permission from copyright holders BC Mental Health and Addiction Services and St Joseph’s Healthcare

[Image of the START Summary Sheet]
What is already known about this topic?

- The START aims to assist practitioners with risk assessment for multiple adverse outcomes including unauthorised leave and substance abuse.
- Its predictive validity for aggression is well-established but it is less clear for relatively infrequent outcomes like unauthorised leave and substance abuse.

What this paper adds?

- START specific risk estimates for unauthorised leave and substance abuse are significant predictors of their respective outcomes.
- The high Negative Predictive Value of the tool suggests it is most useful for screening out low risk patients from more restrictive risk management interventions.
- Researchers should examine risk formulation for these outcomes among clinical teams because their structured judgement outperforms the quantitative element of the START.