International Journal of Forensic Mental Health

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ufmh20

The Short-Term Assessment of Risk and Treatability (START): A Prospective Study of Inpatient Behavior

Nicola S. Gray a, Richard Benson b, Rebecca Craig c, Huw Davies d, Suzanne Fitzgerald c, Phillip Huckle e, Richard Maggs f, John Taylor g, Marianne Trueman h, Tegwyn Williams i & Robert J. Snowden c

a Ty Catrin, Pastoral Cymru, Cardiff, Wales, United Kingdom; and University of Swansea, Swansea, Wales, United Kingdom
b Ty Catrin, Pastoral Cymru, Cardiff, Wales, United Kingdom; and University of Glamorgan, Pontypridd, Wales, United Kingdom
c Cardiff University, Cardiff, Wales, United Kingdom
d Coity Clinic, Abertawe Bro Morgannwg University Health Board, Bridgend, Wales, United Kingdom
e Llanarth Court, Partnerships in Care, Raglan, Wales, United Kingdom
f Cefn Coed Hospital, Abertawe Bro Morgannwg University Health Board, Swansea, Wales, United Kingdom
g Partnerships in Care, Borehamwood, United Kingdom
h Ty Catrin, Pastoral Cymru, Cardiff, Wales, United Kingdom
i Ty Catrin, Pastoral Cymru, Cardiff, Wales, United Kingdom; and Caswell Clinic, Abertawe Bro Morgannwg University Health Board, Bridgend, Wales, United Kingdom

Available online: 07 Dec 2011

To cite this article: Nicola S. Gray, Richard Benson, Rebecca Craig, Huw Davies, Suzanne Fitzgerald, Phillip Huckle, Richard Maggs, John Taylor, Marianne Trueman, Tegwyn Williams & Robert J. Snowden (2011): The Short-Term Assessment of Risk and Treatability (START): A Prospective Study of Inpatient Behavior, International Journal of Forensic Mental Health, 10:4, 305-313

To link to this article: http://dx.doi.org/10.1080/14999013.2011.631692

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
The Short-Term Assessment of Risk and Treatability (START): A Prospective Study of Inpatient Behavior

Nicola S. Gray  
Ty Catrin, Pastoral Cymru, Cardiff, Wales, United Kingdom; and University of Swansea, Swansea, Wales, United Kingdom

Richard Benson  
Ty Catrin, Pastoral Cymru, Cardiff, Wales, United Kingdom; and University of Glamorgan, Pontypridd, Wales, United Kingdom

Rebecca Craig  
Cardiff University, Cardiff, Wales, United Kingdom

Huw Davies  
Coiy Clinic, Abertawe Bro Morgannwg University Health Board, Bridgend, Wales, United Kingdom

Suzanne Fitzgerald  
Cardiff University, Cardiff, Wales, United Kingdom

Phillip Huckle  
Llanarth Court, Partnerships in Care, Raglan, Wales, United Kingdom

Richard Maggs  
Cefn Coed Hospital, Abertawe Bro Morgannwg University Health Board, Swansea, Wales, United Kingdom

John Taylor  
Partnerships in Care, Borehamwood, United Kingdom

Marianne Trueman  
Ty Catrin, Pastoral Cymru, Cardiff, Wales, United Kingdom

Tegwyn Williams  
Ty Catrin, Pastoral Cymru, Cardiff, Wales, United Kingdom; and Caswell Clinic, Abertawe Bro Morgannwg University Health Board, Bridgend, Wales, United Kingdom

Robert J. Snowden  
Cardiff University, Cardiff, Wales, United Kingdom

We thank the staff and patients at Cefn Coed Hospital, The Coity Clinic and Llanarth Court Hospital for their help with this research. The authors have no financial interest in the START.

Address correspondence to Professor Robert J. Snowden, School of Psychology, Cardiff University, Cardiff CF10 3AT, Wales, United Kingdom. E-mail: snowden@cardiff.ac.uk
Structured professional judgment guides (SPJs) have gained acceptance for the prediction of future violence. We conducted a prospective study of 44 psychiatric inpatients with a variety of mental health problems to test whether the Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Middleton, 2004) was able to predict a range of problem behaviors. We obtained outcome behaviors from the nursing record for a period of up to 6 months after the assessment. For all types of behavior tested (violence to others, self-harm, self-neglect, and being victimized), the clinical judgment of risk based on the START was a good predictor. However, the actuarial scores on the Strength and Risk scales of the START were only useful for the prediction of violence. The results provide a strong evidence base for the use of START to predict a range of problem behaviors, and confirms that the START should be used as an adjunct to clinical decision making and not with a blind adherence to the actuarial scores. The difference in efficacy between START used in an actuarial manner and as a SPJ suggests that schemes using other items may prove more effective in guiding the clinician to assess and managing these risks.

The success of the HCR-20 has provided bedrock on which further risk assessments can be built. One issue with the HCR-20 is that it is quite time-consuming to complete and requires experts to administer. Its major use, therefore, seems to be in making decisions about risk management of patients in the shorter term (months), such as decisions relating to release from prison or secure services, child-protection cases, etc. On the other hand, there is also the need to use risk assessments for other tasks, such as in assessing the effects of treatment and management in making day-to-day decisions about a patient (e.g., is it safe to allow the patient to access hospital leave today?) or on admission to emergency services (e.g., acute psychiatric intensive care).

Hence, there are emerging instruments that aim specifically to address such issues. For example, both the Dynamic Appraisal of Situational Aggression (DASA; Olgoff & Daffern, 2006) and the Broset Violence Checklist (BVC: Almvik, Woods & Rasmussen, 2000), aim to provide a brief assessment (5 minutes) of imminent aggression (within the next 24 hours). Both of these instruments have been shown to be useful in making such decisions (Barry-Walsh, Daffern, Duncan, and Olgoff, 2009; Vaaler, Iversen, Morken, Fløvig, Palmstierna, and Linaker, 2011).

While risk of violence to others is of great importance, most clinicians also need to assess (both in the short and long term) other risk areas, such as risk of self-harm, suicide, absconding, or being victimized, etc. Few studies have addressed whether the HCR-20 would have value for these other risks. Gray et al. (2003) did look at the prediction of self-harm/suicide for inpatients and concluded that the HCR-20 was not of great value in this domain.

The Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, and Middleton, 2004) is an SPJ measure whose primary aim is to aid in the management of various types of risk in inpatient settings. START concentrates on dynamic risk factors, aims to look at acute risk, and with the intention of predicting behaviors across a range of risk areas (specifically: Risk to Others, Risk of Self-Harm, Risk of Suicide, Risk of Unauthorised Leave, Risk of Substance Abuse, Risk of Self-Neglect, and Risk of Being victimized). The START is also atypical of most risk assessment instruments in that it assesses both strengths and needs for each of its items.

Here we report on the ability of START to predict some of these risk areas in a sample of UK civil and forensic mentally disordered patients in an inpatient setting.

**METHOD**

**Participants**

Participants were 51 mentally disordered patients who were admitted to one of three psychiatric inpatient units in South Wales; one unit was forensic and two were civil psychiatric units. At the time of assessment, four patients resided in Llanarth Court, a medium-secure forensic hospital for mentally disordered offenders run by the independent sector; 19 patients resided in Cefn Coed Hospital and 28 patients resided in the Coity Clinic, both civil psychiatric hospitals housing patients detained due to issues of risk to themselves or to others.

Of the original 51 patients that we assessed using the START, five were lost to the study as they were discharged from the hospital and we were not able to access their notes in order to score the outcome measures. Two other patients were only resident in the hospital for less than seven days, and so we excluded their data from further analysis. Therefore,
we had complete datasets, with regard to the independent variable (START measures) and the dependent measures (risk behaviors noted in the nursing records), for 44 patients. All further data in this paper refers to these 44 patients.

Diagnoses were retrieved from the patients’ medical records. All diagnoses used ICD-10 criteria. With regard to diagnosis, 66% of patients were diagnosed with schizophrenia related disorder (F20–29), 41% with a mood disorder (F30–39), 14% with a personality disorder (F60–69), 16% from mental and behavioral disorders due to substance misuse (F10–19), 11% with an organic brain disorder (F00–09), and 11% had some other ICD-10 diagnosis relating to Mental and Behavioural disorders (F00–F99) that were not subsumed under the categories already mentioned. Note that these percentages add to far greater than 100% due to co-morbid diagnoses in many patients.

The mean age of the patients at the time of assessment was 40.2 years ($SD = 14.7$, range 20–73). The majority (64%, $n = 28$) was male. All participants described themselves as “white British.”

Design and Procedure

This was a prospective study of inpatient behavior. The independent variables were the START risk and strength scores and the clinical judgment based on the START review. The dependent variables were the problem behaviors on the ward. Because the study utilized a prospective longitudinal design, a control group was not necessary. The Multi-Centre Research Ethics Committee for South West Wales gave ethical approval for the study. We also gained site-specific permissions to carry out this research. We notified participants about the nature and aims of the study both verbally and in writing, and they gave written informed consent. We excluded from the study any patient not considered well enough to give their informed consent by their Responsible Clinician. All patients within the units at the commencement of the study were invited to take part, with no exclusion criteria other than the inability to give informed consent. No patient was actually excluded on this basis.

Assessments involved access to file-based information and an interview with the patient. Assessors had access to the information available within the multidisciplinary clinical records held within the units, including psychometric tests when administered and collateral records. This aimed to replicate the situation many clinicians face when using the START in an acute setting where many detailed psychological reports, etc., are not yet available.

All of the assessors were trained by a Consultant Clinical and Forensic Psychologist accredited by the British Psychological Society and who has much experience in the administration of SPJs, as well as in training professionals on other SPJs. The raters were a senior nurse, and two psychology assistants.

Measures

**Short-Term Assessment of Risk and Treatability (START)**

START is a 20-item structured professional judgment (SPJ) scheme. For each item, the clinician must judge on a 3-point scale as to whether this risk factor is present (2), partially present (1), or not present (0). They also judge if this item is a “strength” for this person using the same numerical scale, but slightly different quantitative judgments meaning that strengths are not just the mirror image of risk factors. After completion of the 20 items each person has a score from 0–40 for the Risk scale, and from 0–40 for the Strength scale.

Nickolls, Brink, Demarais, Webster, and Martin (2006) have shown that these have good interrater reliability (ICC = 0.87), and good internal consistency (0.80–0.97). However, as with other SPJ schemes, START should not be used merely in an actuarial manner (i.e., by adding up the scores), but should provide the basis for making a clinical judgment based on the pattern of risks and strengths. Therefore, at the end of assessment, the assessor made a clinical judgment (on a 5-point scale from 1 = very low, 2 = low, 3 = medium, 4 = high, and 5 = very high; this differs a little from the scheme in the START manual which offers only a 3-point scale) as to the likelihood of each particular behavior (e.g., self-harm) occurring within the next eight weeks. A 5-point scale was utilized as we have used this system in previous published research on other risk assessment instruments and to give a finer analysis to the risk judgments.

**Aggression Vulnerability Questionnaire (AVQ)**

The Aggression Vulnerability Questionnaire (AVQ) is a coding scheme to quantify inpatient incidences recorded in nursing records, or similar records, designed to log any untoward events within the unit (see Gray et al., 2003). For each entry in the record the assessor decides as to which type of incident this belongs (from the choices of: Verbal Aggression, Aggression against Property, Physical Aggression, Self-Harm/Suicide, Self-Neglect, Victimization, Sexual Vulnerability, Absconding Behavior, or Non-compliant Behavior). Within the chosen category, the assessor then chooses the level of severity of the incident that is hierarchically organized. So, for example, within the category of self-harm/suicide there are eight possible levels ranging from the least serious (level 1 = Verbally threatening violence to self), through more serious (level 4 = bangs head, inflicts minor cuts, bruises or burns, etc. to self) to the most serious (level 8 = completed suicide).

Data Analysis

Our predictor measures consisted of the START scores from the Strengths and Risk scales, and five SPJ judgments in the domains of harm to others, self-harm, suicide, self-neglect and victimization. Our outcome measures consisted of a
severity score, and category type, using the AVS for each incident recorded in the nursing records. As patients often left before the end of our follow-up period (see Results), we calculated an incident rate (irrespective of severity), by dividing the number of incidents in each category by the number of days over which we had followed the patient. For convenience, we multiplied these numbers by 100 to express these as incident rate per 100 days. Some analyses have also used the severity rating to limit the analyses to only certain levels of severity (see Results), and then calculated this rate in the same manner.

For each risk area we have performed two analyses. Our primary analysis is based on signal detection theory, which aims to see how well the predictor variable is able to distinguish between those that showed a particular behavior and those that did not. To do this, we had to produce a binary outcome measure (behavior was present or not), and then construct the Receiver Operating Characteristic (ROC; a plot of proportion of correct predictions against proportion of false alarms; see Figure 1). Such analyses are used in many areas of medicine as the appropriate measure of an instrument’s predictive accuracy (see Snowden & Gray, 2010). To quantify the ROC we calculated the Area Under the Curve (AUC). The AUC can be interpreted as the probability that an individual selected at random from among those with a dependent characteristic (e.g., violent behavior) will have a higher score on the independent index (e.g., START score) than an individual selected at random from those without the dependent characteristic.

The ROC analysis loses some aspects of the data, as each person’s behavior is reduced to a binary outcome (behavior present or not), hence data such as the frequency of such incidents, severity of incident, or the time to first incident, are lost. We, therefore, supplemented our ROC analysis by correlating the incident rate against the predictor variables.

Analysis of the outcome variables showed very high levels of non-normality in the data. There were no obvious transforms that could correct for this. Therefore, all analyses used non-parametric statistics (AUC for the ROC, and Spearman’s ρ for the correlations). We also present means (as well as medians) for further information. As we are predicting that greater scores will be predictive of more target behaviors all tests were one-tailed (note that for the Strengths scale this prediction is inverted as we predict greater strengths will be predicted of fewer problem behaviors).

RESULTS

Overview

As expected, the majority of the patients (n = 26; 59%) did not remain within the clinic for six months after we assessed them. The median length of stay/follow-up period was 154 days (M = 114, SD = 73.7).

Risk Scores

Actuarial scores. Figure 2 depicts the frequency scores for the strength and risk measures and Table 1 provides their descriptive statistics. The strength scores are not normally distributed (Skewness = 0.80, SE = .33), with a heavy preponderance for low scores. The risk scores are also not normally distributed (Skewness = –0.74, SE = .33), with a preponderance of high scores. As might be expected, there was a strong inverse correlation between the strength and risk scores (ρ = –.77, p < .001; see Table 1).

Structured Professional Judgment. START was designed as an aid to the professional, rather than just a method for producing an actuarial measure of risks. Figure 3 shows the distribution of the clinically predicted level of risk for each of the risk areas that we tested. The diagrams illustrate that within our population we could discern a broad range of apparent risk. For each risk area, at least one individual populates both the upper and lower risk categories. For most risk areas there appears to be some skew, with greater frequencies of individuals in the lower risk categories.

Association Between Predictor Variables

Table 1 presents the relationships between the various predictor variables. As mentioned previously, there was a strong inverse relationship between the Strength and Risk scales. The Strength scale was positively predictive of the SPJ clinical rating for most risk areas. However, it is notable that the correlation with the SPJ for self-harm and for suicide was actually positive, and significantly so in the case of the suicide SPJ. Hence, those judged to have many strengths were
more likely to be predicted as a suicide risk than those with few strengths! This, perhaps surprising, result may reflect the clinical finding that those people with greater strengths (such as intelligence; Webb, Långström, Runeson, Lichtenstein, & Fazel, 2011) are at greater risk of completed suicide.

The Risk scale was positively related to the SPJ risk of violence to others, self-neglect, and of being victimized by others. However, the Risk scale was not significantly correlated with the SPJ for self-harm or for suicide.

Table 1 also documents the relationships between the SPJs. Most of these associations between the SPJs were small and statistically insignificant. This suggests that the raters are not merely deciding on one level of risk for an individual and then applying this to each area of risk. The exception to this lack of association was the strong association between SPJs for self-harm and suicide. This may reflect the clinical reality of how difficult it can be to accurately and confidently differentiate between an episode of deliberate self-harm and a genuine suicide attempt.

**Aggression**

Of the 44 patients, 11 (25%) committed at least one act of physical violence towards another person. The distribution was highly skewed, with most individuals having a mean rate of 0 acts per 100 days, and the most being 11 incidents per 100 days ($Mdn = 0; M = 0.93, SD = 2.46$). Table 2 illustrates the predictive properties of START for the prediction of violence to others. The Strength scale was negatively associated to the incident rate of violence. In line with this, the AUC was significantly less than 0.5. The Strength scale produced a large effect size (Rice & Harris, 2005). The Risk scale was also significantly predictive of future violence with a medium effect size. The SPJ predictive accuracy just failed to reach statistical significance ($p < .08$), though produced a medium effect size.

We also examined acts of verbal aggression, using the SPJ for the prediction of physical violence. Of the 44 patients, 23 (52%) committed at least one act of verbal aggression towards another person. Again, the distribution was highly skewed, with most individuals having a mean rate of 0 acts per 100 days, and the most being 32 incidents per 100 days ($Mdn = 2.03, M = 3.79, SD = 5.24$). Table 2 also illustrates the predictive properties of START for the prediction of verbal aggression. All scales, including the SPJ, were significant predictors of this aggression, and all had very similar effect sizes.

**TABLE 1**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td></td>
<td>–</td>
<td>–.77**</td>
<td>–.35**</td>
<td>.26</td>
<td>.36**</td>
<td>–.35**</td>
</tr>
<tr>
<td>R</td>
<td>–</td>
<td></td>
<td>–</td>
<td>–.33*</td>
<td>–.02</td>
<td>–.23</td>
<td>.35**</td>
</tr>
<tr>
<td>Violence</td>
<td>–</td>
<td></td>
<td></td>
<td>–.08</td>
<td>–.12</td>
<td>.15</td>
<td>–.02</td>
</tr>
<tr>
<td>Self-harm</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td>–.64**</td>
<td>–.10</td>
<td>–.09</td>
</tr>
<tr>
<td>Suicide</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>–.12</td>
<td>–.14</td>
</tr>
<tr>
<td>Neglect</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>–.22</td>
</tr>
<tr>
<td>Victim</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05, **p < .01 (both one-tailed).
Self-Harm

Of the 44 patients, 14 (32%) committed at least one act of self-harm. The distribution was highly skewed, with most individuals having a mean rate of 0 incidents per 100 days, and the most being 19 incidents per 100 days ($Mdn = 0$, $M = 2.27$, $SD = 5.07$). Table 3 illustrates the predictive properties of START for the prediction of self-harm. None of the scales, when used as an actuarial measure, predicted self-harming behaviors. Indeed, the correlations (though non-significant) were in the opposite direction to predictions (and for those found for risk of violence to others). Despite this, the SPJ judgments of future self-harm were strongly correlated with self-harm outcomes and were highly accurate, producing very large effect sizes.

We also intended to perform a separate analysis that looked at suicide attempts. Our AVS attempts to classify self-injurious behaviors as either self-harm or as a suicide attempt (in the former case there is no intent to die, while in the latter there is such intent; for a discussion see Claes & Vandereycken, 2007). However, in the follow-up time period, none of the instances of self-harm were judged to be suicide attempts.

Self-Neglect

Of the 44 patients, 16 (36%) committed at least one act of self-neglect. The distribution was highly skewed, with most individuals having a mean rate of 0 acts per 100 days, and the most being 6 incidents per 100 days ($Mdn = 0$, $M = 1.38$, $SD = 3.68$). Table 3 illustrates the predictive properties of START for the prediction of self-neglect. None of the scales, when used as an actuarial measure, predicted self-neglecting behaviors. Indeed, the correlations (though non-significant) were in the opposite direction to predictions (and for those found for risk of violence to others). Despite this, the SPJ judgments of future self-neglect were strongly correlated with self-neglect outcomes and were highly accurate, producing very large effect sizes.

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Predictive Accuracy (Spearman’s $\rho$ and AUC) of START With Respect to Violence to Others and Verbal Aggression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Violence to Others</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
</tr>
<tr>
<td>S</td>
<td>-.42**</td>
</tr>
<tr>
<td>R</td>
<td>.28*</td>
</tr>
<tr>
<td>SPJ</td>
<td>.23</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01 (both one-tailed).*
Predictive Accuracy (Spearman’s ρ and AUC) of START With Respect to Self-Harm

<table>
<thead>
<tr>
<th></th>
<th>ρ</th>
<th>AUC</th>
<th>(SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>.22</td>
<td>.61</td>
<td>(.10)</td>
</tr>
<tr>
<td>R</td>
<td>-.02</td>
<td>.48</td>
<td>(.10)</td>
</tr>
<tr>
<td>SPJ</td>
<td>.63**</td>
<td>.86**</td>
<td>(.06)</td>
</tr>
</tbody>
</table>

* p < .05, **p < .01 (both one-tailed).

Predictive Accuracy (Spearman’s ρ and AUC) of START With Respect to Victimization

<table>
<thead>
<tr>
<th></th>
<th>ρ</th>
<th>AUC</th>
<th>(SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>-.02</td>
<td>.47</td>
<td>(.10)</td>
</tr>
<tr>
<td>R</td>
<td>.14</td>
<td>.60</td>
<td>(.09)</td>
</tr>
<tr>
<td>SPJ</td>
<td>.24</td>
<td>.67*</td>
<td>(.09)</td>
</tr>
</tbody>
</table>

* p < .05, **p < .01 (both one-tailed).

As hoped, our sample showed a wide range of START scores and of SPJ categories on the START and hence provided a chance to examine the predictive efficacy of these scores and judgments. Our results show that we were able to predict a wide range of problem behaviors using the structured professional judgment derived from the START. Indeed, in some cases (e.g., the prediction of acts of self-harm) the prediction accuracy was very high. Therefore, our results provide an evidence base for the use the SPJ component of the START in inpatient settings.

The items of the START are scored as to whether this person has a risk (or vulnerability) based on each item, and whether they have a strength on each item (note that Webster et al., 2004 believe that strengths need not be the obverse of risk factors; see also Durrlak, 1998). By adding these scores we obtained two actuarial scores based on the Strength scale and the Risk scale. As expected, these actuarial scores were also good predictors of violent behavior. Hence, the items of the START seem appropriate to help the clinician make this judgment about violence, and this is supported by the correlations between these actuarial scales and the SPJ for violence.

For the prediction of self-harm the actuarial scores were not significantly correlated in a predictable manner with the outcome. Indeed, the insignificant correlations were actually in the opposite direction to that expected (i.e., those high on Strengths were more likely to commit such behaviors). Further research with a more powerful design is needed to see if this suggestive trend is replicable. We note that Braithwaite, Charrette, Crocker, and Reyes (2010) also found that the START score was not predictive of self-harm or suicidality. Thus, at best, the scores obtained from the items of the START are not useful in predicting who would self-harm. Nevertheless, the SPJ based on the START was an excellent predictor of self-harming behaviors. Clearly, the raters must not be placing much weight upon the overall number of risk or strengths of the START items when making their judgments. Indeed, our evidence seems to suggest that the SPJs were inversely related to scores on these scales in some cases. This raises two points. First, the items of the START do not appear to be good predictors of self-injurious behaviors. Previous research by Gray et al. (2003) has shown that another scheme designed to predict violence (the HCR-20) was not predictive of self-injurious behavior. Hence, it appears that the risk (and protective) factors for these two types of behavior are sufficiently different that it has not yet been
possible to produce a scheme that aids in the prediction of both of these behaviors. It may be that a separate SPJ scheme using different risk factors may prove more efficacious in predicting self-injurious behaviors.2

The actuarial scores of the START were only weakly predictive of self-neglecting behaviors, yet the SPJ was a strong predictor. The arguments we have just made in the case of self-injurious behaviors therefore also apply to this other risk behavior. We note that Braithwaite et al. (2010) also show that START scores in their sample of inpatients was not predictive of self-neglecting behaviors.

Finally, the actuarial scores of the START were not predictive of being victimized, and the SPJ measure was only a weak indicator of being victimized. Again, Braithwaite et al. (2010) show that START score in their sample of inpatients was not predictive of being victimized.

For most risk areas (with the exception of physical violence) there appears to be a pattern of results whereby the scores on the START when used in an actuarial manner are not predictive of future behavior, but when used as a guide to the SPJ the START is highly effective. This may be surprising, but the nature of an SPJ measure is that the assessor makes clinical decisions based on a careful review of items, and not a blind adherence to the scores on the scale. Hence, a person maybe thought of as a high risk even though there is only one item present (e.g., a strong command hallucination to kill themselves), or a low risk because of protective factors (e.g., a strong religious belief that they must not kill themselves.). Hence, the actuarial score on the instrument need not, and should not, be deterministic of the SPJ measure. Indeed, Table 1 shows that the correlations between the actuarial START scores and the SPJ measure for violence are only modest. However, for the prediction of self-harm we note that the scores used in an actuarial manner appear to run counter to the SPJ measure. This pattern poses the question of just how the SPJ emerges from the rater’s consideration of the items. Clearly, the assessors must have been able to weight certain information as highly valuable in making these predictions. For example, they may have employed a “like predicts like” strategy (e.g., Sandford, Gray, Taylor, & Snowden, 2011), for example, self-harming in the past is the best indicator of self-harming in the future - in making these judgments even if this information did not appear within the START itself. We also note that Braithwaite et al. (2010) developed an “optimized” START score that used only the individual items that were effective for each type of risk. This optimized value was then predictive for risks that the overall START was not (e.g., risk of self-harm). In theory, our assessors could have been able to do a similar sort of mental weighting of items, but this seems unlikely given that they did not have the information of which items would be the most useful. The exact nature of how the assessor goes from the information about individual items to a global SPJ of risk of a particular behavior remains unknown, and an area that needs future investigation.

Limitations

The main limitation of the study is the small sample size and this will be likely to lead to Type II errors (not finding that the measure was predictive of the behavior when it really is) rather than Type I errors. Hence, our findings that START was effective even in this small sample is most encouraging. Further, given this small sample size, we were unable to split the sample into particular groups of interest (e.g., different levels of clinical security, gender, diagnosis) and the data presented is drawn from a mixture of both forensic and non-forensic patients. Of course, many clinicians have to deal with such “mixed populations” and therefore our demonstration that the START is effective in this population is, we hope, of value. Nevertheless, further research is needed to see how this instrument performs with particular populations given that others have shown that the efficacy of some instruments can vary with both factors as diagnosis (Gray, Taylor, & Snowden, 2011).

We note that our outcome measure was scored from file information only (nursing records), which may not be a faithful record of all incidents that took place. It may be that certain behaviors are missed or not recorded in these nursing records.

We would like to suggest that our finding that the professional judgment based on the START outperformed that of the START used in an actuarial manner was good evidence of the efficacy of the START method. However, we did not include in our protocol any unstructured clinical judgment of these risks (blind to the START analyses) to which to compare the SPJs. It is possible that (unstructured) clinical judgment alone may also be as effective in the prediction of these behaviors. Previous research, however, suggests that unstructured clinical judgment of many of these behaviors is poor (Gustafson, Greist, Stauss, Erdman, & Laughren, 1977; Monahan, 1981).

Another limitation of our study is that we did not test all the areas of risk that are covered in the START. Notably, we did not collect data relating to risk of unauthorized leave or risk of substance abuse due to a very low expected base rate in the institutions included.

Our sample did not include all the patients in these hospitals. For ethical reasons, we were unable to collect data on patients that were regarded as unable to give informed consent, and those that did not want to give consent. This, therefore, may well produce biases in the sample. As our ethical procedures did not allow us to have information about these patients who refused to take part, we were unable to

---

2The present results have led us to develop such an SPJ for suicide and attempted suicide called the Risk of Suicide Protocol (RoSP) that we are now using and researching. A draft version is available from the corresponding author.
compare the obtained sample to those that we were unable to test. It may well be that these people who refuse to consent to participate may have been higher risk individuals or those with elevated rates of psychosis, learning disabilities, and personality disorders compared to the consenting sample. Clearly, given the possible importance of these factors, and that clinicians are still called on to perform risk assessments on people who are not able, or not willing, to consent to being assessed, it would be most valuable to have a study of risk assessment instrument(s) which overcame the ethical problems of including these people.

Conclusions

We conclude that the SPJ based on START is a valuable method for predicting a range of future risk behaviors in both forensic and civil psychiatric settings. However, we found that an actuarial analysis of scores based on the individual items of the START were not predictive of most of these behaviors. Hence, the monitoring of such scores as a guide to treatment outcome, or as a means to demonstrate changing risk, must be brought into question. Rather, the clinical professional judgment resulting from a careful review of these risk factors is needed to demonstrate such clinical outcomes.

REFERENCES


