



Role of assessment components and recent adverse outcomes in risk estimation and prediction: Use of the Short Term Assessment of Risk and Treatability (START) in an adult secure inpatient mental health service



Laura E. O'Shea^a, Geoffrey L. Dickens^{b,*}

^a St Andrew's Healthcare, Northampton NN1 5DG, United Kingdom

^b School of Social and Health Sciences, Abertay University, Bell Street, Dundee DD1 1HG, United Kingdom

ARTICLE INFO

Article history:

Received 8 July 2015

Received in revised form

28 March 2016

Accepted 21 April 2016

Available online 22 April 2016

Keywords:

Risk assessment

Decision making

Violence

Deliberate self-harm

Victimisation

Self-neglect

ABSTRACT

The Short Term Assessment of Risk and Treatability is a structured judgement tool used to inform risk estimation for multiple adverse outcomes. In research, risk estimates outperform the tool's strength and vulnerability scales for violence prediction. Little is known about what its component parts contribute to the assignment of risk estimates and how those estimates fare in prediction of non-violent adverse outcomes compared with the structured components. START assessment and outcomes data from a secure mental health service ($N=84$) was collected. Binomial and multinomial regression analyses determined the contribution of selected elements of the START structured domain and recent adverse risk events to risk estimates and outcomes prediction for violence, self-harm/suicidality, victimisation, and self-neglect. START vulnerabilities and lifetime history of violence, predicted the violence risk estimate; self-harm and victimisation estimates were predicted only by corresponding recent adverse events. Recent adverse events uniquely predicted all corresponding outcomes, with the exception of self-neglect which was predicted by the strength scale. Only for victimisation did the risk estimate outperform prediction based on the START components and recent adverse events. In the absence of recent corresponding risk behaviour, restrictions imposed on the basis of START-informed risk estimates could be unwarranted and may be unethical.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Structured professional judgment (SPJ) is the gold standard violence risk assessment technique in mental health care (National Institute for Mental Health in England, 2004). Such schemes promote risk-related assessment and decision-making through the systematic consideration of a domain of structured components, including empirically derived risk factors, and a clinical professional judgement domain comprising components of case-specific knowledge and evaluator discretion (Hart and Logan, 2011). One aim of SPJ schemes is to facilitate the formulation of risk estimates or summary judgments, usually articulated in terms of low, moderate or high risk, based on a synthesis of these two domains. The importance of clinicians' risk estimates in legal and clinical decision-making has been extensively documented (Dvoskin and

Heilbrun, 2001; Skeem et al., 2005; Buchanan, 2013), as have the consequences of inaccurate assessments (Skeem et al., 2005; de Ruiter and Nicholls, 2011; Large et al., 2011; Szmukler and Rose, 2013). However, the vast majority of research regarding risk assessment has been prescriptive about the factors that clinicians should consider when assessing risk; relatively little descriptive research has been conducted to investigate what clinicians actually do when assessing risk (Elbogen, 2002).

The Short-Term Assessment of Risk and Treatability (START; Webster et al., 2009) is a recently developed tool which attempts to extend the SPJ approach. Our schematic of the START assessment process using violence as an example adverse outcome is presented in Fig. 1. Importantly, the START aims to guide risk formulation for a range of the adverse outcomes that are faced by people with mental disorder in addition to violence, namely self-harm, suicide, self-neglect, victimisation, substance abuse and unauthorised leave. The START manual instructs raters to assign a specific risk estimate (or SRE), categorised as high, moderate, or low, for each of the seven potential risk outcomes. In contrast to deficits-based approaches to risk assessment, which focus largely

* Corresponding author.

E-mail addresses: leoshea@standrew.co.uk (L.E. O'Shea), g.dickens@abertay.ac.uk (G.L. Dickens).

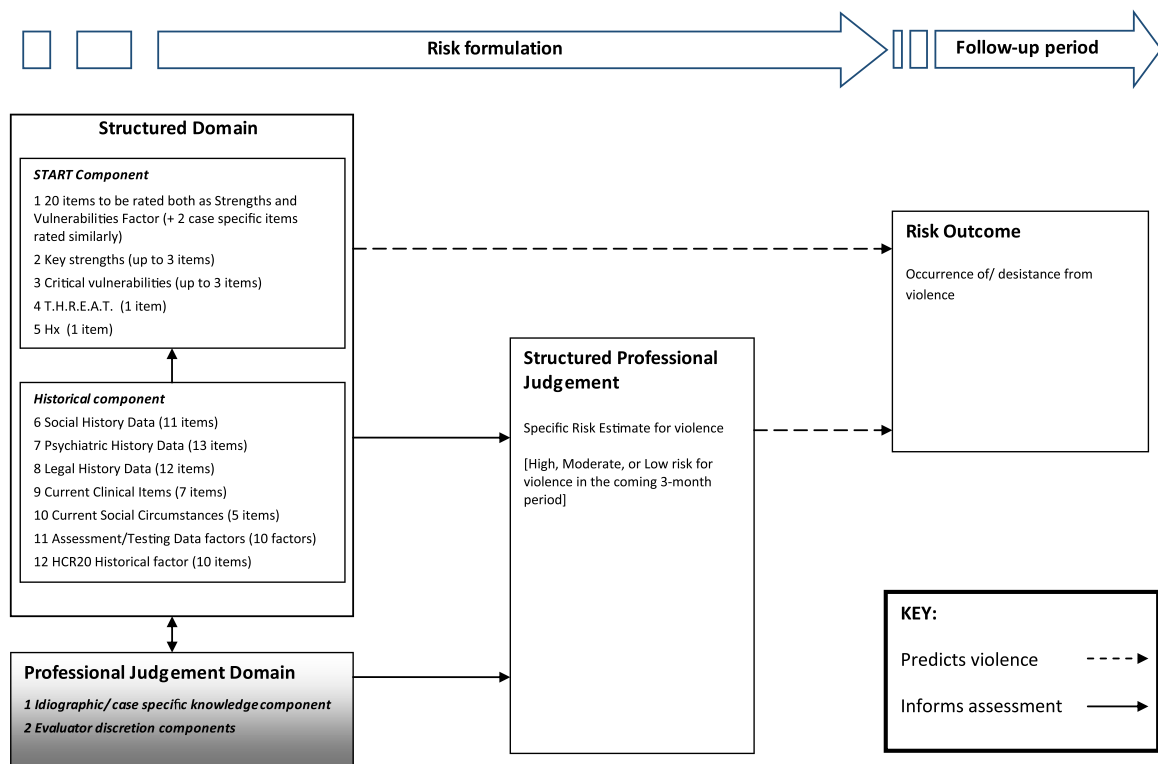


Fig. 1. Conceptual model of START risk assessment process using outcome violence as an example.

on risk factors held by the person under assessment (Rogers, 2000), the structured domain of the START formulation requires consideration of 20 empirically derived items (plus up to two further case specific items) both as potential protective factors (or 'strengths') and as risk factors (or 'vulnerabilities'). Additional innovations include a suggestion to identify any key strength and critical vulnerability items, to highlight potentially imminent and/or severe violent and self-harm/suicidal outcomes (i.e., threats of harm that are real, enactable, acute and targeted [T.H.R.E.A.T.]), and to record the presence of previous lifetime history (Hx) for the individual under assessment of adverse events corresponding to each adverse risk outcome. In addition to these 48 START component items, the START manual suggests consideration of historical information comprising 68 items related to social, psychiatric, and legal history, current clinical and social circumstances, and assessment/testing data (including the 10 historical factors of the HCR-20). These items might inform rating of the 20 START component items and how much weight they are given in the final SRE; they might also independently contribute to the formulation of each SRE. Finally, in common with other SPJ tools, it is anticipated that raters will use their idiographic clinical knowledge of the individual patient and their own discretion (Guy et al., 2012) to aid risk formulation. Formulation of each of the seven START SREs therefore requires consideration of up to 116 specified items per patient per assessment. Repeated assessment is recommended at 3-monthly intervals.

The ultimate purpose of the START is to inform therapeutic endeavours including clear communication of the nature, probability, type, imminence, and severity of risk and thus the resulting SRE should be expected to *predict* actual outcomes in order that appropriate management measures can be implemented to mitigate that risk. Additionally, we expect that components from the structured domain of the START assessment will *predict* actual outcomes since they have presumably been selected by the START authors precisely for their relationship with outcomes as risk factors. These components may have been considered both on

their own merit and in the light of components from the professional judgement domain. The START component of the structured domain may also have been formulated in the light of the historical component. While it may not be possible to precisely define the effect on predictive ability of assessment components in the professional judgement domain, we can reasonably infer that it is the intention that those components will improve the formulation such that predictive value of the risk estimate will be at least the equal of that achieved by the components of the structured domain. Since we cannot discount that structured domain components have themselves been influenced by those in the professional judgement domain then it is too strict a test to expect that the risk estimate will significantly improve prediction. Since the START is designed not only to facilitate risk assessment but to inform and improve risk management then our claim that certain elements should improve prediction may seem counterintuitive since improvements in management should lead to reductions in adverse outcomes and a concomitantly poorer predictive performance. However, there is good evidence to suggest that the clinician-rated START risk estimates are predictive of aggression, self-harm, and victimisation in clinical practice (O'Shea et al., 2016).

Despite the apparent sophistication and comprehensive nature of the START assessment model there has been little research into the predictive value of its components. Research has revealed that START strength and vulnerability total scores are significantly predictive of violent and aggressive outcomes (O'Shea and Dickens, 2014); however, SREs for the same outcomes produce larger effect sizes, and SREs for self-harm and victimisation predict their associated outcomes whilst scale scores do not (O'Shea and Dickens, 2014; O'Shea et al., 2016). Further, only moderate correlations have been found between the total START Vulnerability or Strength scores and the SREs (O'Shea et al., 2016). In rarer cases, the risk formulation process results in SREs that are poorer predictors of outcome than the empirically derived risk and protective factor scores. For example, a recent study found that the START SRE for self-neglect did not predict its corresponding outcome

despite predictive validity achieved by the START Strength scores (O'Shea et al., 2016). Thus, clinicians appear to be arriving at SREs which differ from those based solely on scale scores and are not simply transforming scores into overall risk estimates; this would largely be expected given that the presence of empirical risk factors is intended to be supplemented by the wide range of other factors described above. Currently, the weight of evidence suggests that SREs improve prediction over scale scores; nevertheless, even worsening prediction would suggest that important considerations are being made by raters during the formulation process. Therefore, it would be valuable to identify which risk formulation components play an important role in clinicians' determination of START SREs; and especially important to identify those components that are considered when SREs either improve or worsen prediction since this would inform the weight such components are afforded in the overall formulation. Given also that findings from feasibility studies have revealed that mental health professionals experience some difficulty in completing START SREs (Doyle et al., 2008; Crocker et al., 2011) there is a practical need for evidence to inform the content of guidance and training for mental health professionals tasked with completing the START. We hypothesised, in line with our conceptualisation of the START assessment process, that SREs would be predicted by structured assessment components; and that outcomes would be predicted by both SREs and structured assessment components.

2. Methods

2.1. Study model and conceptualisation

Given the large number of component items included in the START assessment process we have developed a streamlined version to underpin this exploratory study (see Fig. 2). Modification involved a pragmatic consideration of the available data, of the feasibility of converting START items/factors into meaningful operational variables, and the need to limit the number of variables under investigation to a parsimonious level given the sample size involved. While our operational study model reflects our overall conceptual model of START assessment it comprises only five predictor variables (see Fig. 2) and does not address possible interactions. We have added an additional component item into the model, namely recent (past 3-month) engagement in each risk outcome of interest. There is no such explicit item in the START scheme; however, from a theoretical point of view, one might reasonably expect past behaviour to predict future behaviour where prediction is

for a short interval (Oullette and Wood, 1998). Sample size for the current study was predicated on the 'rule of ten' cases per variable (Peduzzi et al., 1996).

2.2. Participants

Potential participants were all consecutive admissions between May 2011 and July 2012 to St Andrew's, a large charitable provider of secure mental health care at four sites in England. In order to be eligible, patients must have had at least one START and HCR-20 risk assessment completed and been resident in the service for at least three months before and after START assessment.

2.3. Procedure

A pseudo-prospective design was used; START and HCR-20 assessments were completed by each patients' clinical team as part of routine clinical care and risk incidents were recorded on each shift by a qualified member of staff. The first START assessment and the HCR-20 assessment that preceded the START were extracted from patient records, along with demographic and clinical data, and information about risk incidents involving violence, self-harm/suicidality, victimisation, and self-neglect for the three months preceding and following START assessment. We did not include substance misuse or unauthorised leave since we have previously found both to have very low base rates (O'Shea and Dickens, 2015b).

2.4. Measures

2.4.1. START assessment

The START is intended to assess the likelihood of seven adverse outcomes occurring over a maximum of three months: violence to others, self-harm, suicide, substance abuse, victimization, self-neglect, and unauthorised absence. It comprises 20 dynamic items that are scored, in the current study setting, by a multi-disciplinary team of at least three practitioners from various professions on two 3-point scales, ranging from 0 (no/minimal) to 2 (high); once regarding risk factors or "vulnerabilities" and once in terms of protective factors or "strengths". Raters then consider any additional case-specific factors, indicate key strengths and critical vulnerabilities among the 20 items, and record the presence of a lifetime history of any of the risk outcomes. Teams of raters then make a specific risk estimate (SRE; low, moderate, or high) regarding the likelihood of each outcome occurring. There is limited guidance regarding the formulation of the SREs, but raters are advised to consider items detailed in Fig. 1 (Webster et al., 2009).

2.4.2. HCR-20 assessment

Version 2 of the HCR-20 (Webster et al., 1997) was used as version 3 (Douglas et al., 2013) was not in use in the study setting at the time of data collection. The HCR-20 comprises 20 items, 10 historical (H10), 5 clinical (C5), and 5 risk-management (R5). Each item is rated on a 3-point scale as not present (0/No), partially present (1/Maybe), or definitely present (2/Yes). In the current setting rating is

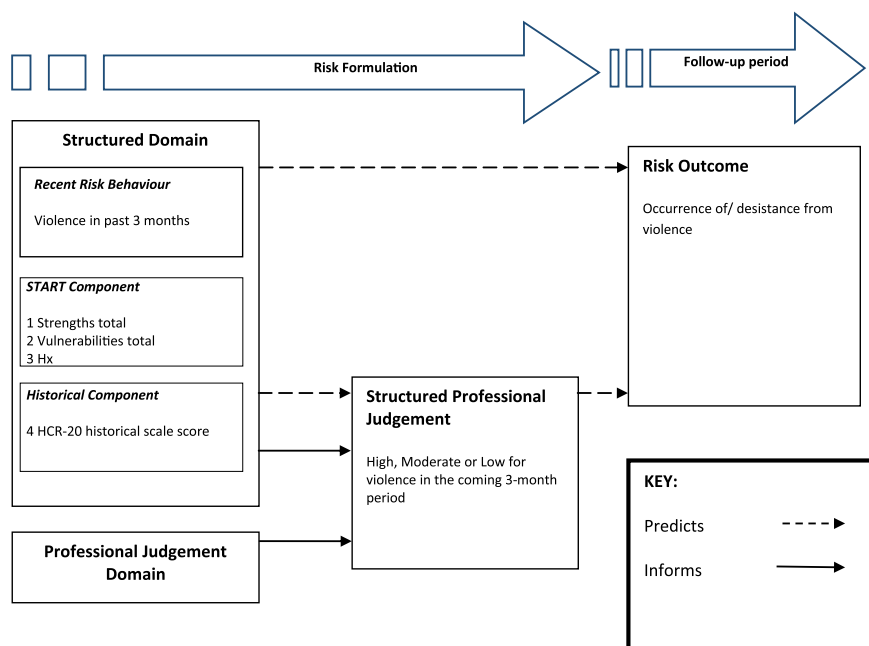


Fig. 2. Modified conceptual model for study using outcome violence as an example.

conducted by registered psychologists or graduate psychology assistants under their supervision. Only the H10 items were included for the purpose of the current study; the psychopathy item was omitted as the majority of patients had not had a formal psychopathy assessment. Scores were prorated in accordance with guidance in the HCR-20 manual (Webster et al., 1997).

2.4.3. Demographic and clinical data

Age, gender, security level, legal status, admission and discharge dates, self-reported ethnicity and ICD-10 (World Health Organisation, 1992) psychiatric diagnoses were extracted from patient records.

2.4.4. Risk Behaviours

Hospital policy dictates that a free text electronic progress note is entered by a qualified member of the patients' multidisciplinary team for each patient, on a per shift basis. Staff can electronically flag the note if any of a range of risk outcomes has occurred. As part of a previous study we (LEO, GLD) collated all outcomes flagged as "Aggression – Physical", "Aggression – Verbal", "Absconding", "Self-harm/Suicide", "Self-neglect", "Substance Misuse", and "Vulnerability, and coded incidents occurring after assessment; inter-rater reliability was in the excellent range ($\kappa=0.83$ – 1.00 ; O'Shea et al., 2016). Incidents were coded using the START Outcome Scale (SOS; Nicholls et al., 2007) which comprises 12 outcome categories, rated on a criterion-referenced severity scale of 0 (outcome absent) to 4 (most severe): verbal aggression, aggression against property, physical aggression against others, sexual aggression, self-harm, suicide ideation and planning, suicide behaviours, self-neglect, unauthorised leave, substance abuse, being victimised, and stalking. We did not code entries separately for sexual aggression and stalking since the START is not intended to predict these outcomes; Aggression against property and physical aggression against others were analysed as a single outcome (physical aggression) as were self-harm, suicide ideation and planning, and suicide behaviours (self-harm/suicide); additionally, we combined incidents of physical aggression and verbal aggression to form an "Any Aggression" category (see O'Shea et al. (2016)). Incidents occurring prior to assessment were coded in the same way by LEO and an undergraduate psychology student who received training in rating. Inter-rater reliability was again measured on a sample of progress notes for 20 patients per outcome. Presence of adverse events corresponding to each risk outcome in the three months before and after assessment was dichotomised as present or absent.

2.5. Data analysis

Descriptive statistics were calculated to examine sample characteristics, the distribution of START scores and SREs, Historical HCR-20 scores, and the incidence of risk outcomes before and after START assessment. Multinomial logistic regression was used to examine the contribution of recent behaviour and the various START and HCR-20 components to the SREs; specific risk estimates were entered as dependent variables in turn, with the low risk group defined as the reference category. START Strength scores, START Vulnerability scores and H10 scores were entered as covariates, and the corresponding flagged risk history from the START assessment, and the presence of risk behaviours in the three months preceding START assessment were entered as factors (reference category = absence of the history/behaviour). Cox and Snell's and Nagelkerke's R^2 were reported as an indication of how well the model fits the data; Cox and Snell's R^2 is more commonly used, but has a maximum value of less than 1, making it difficult to interpret. Nagelkerke's R^2 (Nagelkerke, 1991) is more similar to R^2 values generated from linear regression (Eberhart and Hammen, 2006), which can be explained as the proportion of the variance explained by the predictors in the model (Peng et al., 2002). Binary logistic regression was then used to examine the ability of the component parts, recent adverse outcomes, and SREs to predict adverse outcomes in the following three months (for categorical variables, reference category = absence of the history/behaviour). The START components and recent adverse outcomes were entered in step 1 and the SREs were entered in step 2 to investigate whether they had incremental validity over the component parts alone. All analyses were conducted using PASW Statistics version 18 for Windows (SPSS Inc, Chicago, US).

3. Results

3.1. Inter-rater reliability

Inter-rater reliability for the coding of behaviour in the three months prior to START assessment was in the good to excellent range; Cohen's kappa ranged from 0.64 to 1.00, mean 0.89.

3.2. Sample characteristics

There were 84 patients who met the inclusion criteria; mean age at assessment was 34.15 years ($SD=14.18$). Nearly three quarters of the sample were male ($n=61$, 72.6%); approximately half ($n=39$, 46.4%) were Caucasian, 3 (3.6%) were of Asian background, three (3.6%) were Black, two (2.4%) were of mixed ethnic background and the remainder ($n=37$, 44%) had no recorded ethnicity. Most common ICD-10 psychiatric diagnoses were "Schizophrenia, schizotypal and delusional disorders" (F20–F29; $n=35$, 41.7%), "Disorders of psychological development" (F80–F89; $n=27$, 32.1%), "Disorders of adult personality and behaviour" (F60–F69; $n=25$, 29.8%), and "Mental retardation" (F70–F79; $n=16$, 19%); hereafter referred to as intellectual disability. There were 52 (61.9%) patients admitted under forensic sections of the Mental Health Act (1983, amended in 2007), 31 (36.9%) admitted under civil sections, and one (1.2%) informal admission; 49 patients (58.3%) were resident in medium secure wards and 35 (41.7%) were in low secure wards. The mean time between admission and the first START assessment was 245 days ($SD=109.8$; range 39.9–488.9); the mean time between START assessment and HCR-20 assessment was 117.2 days ($SD=92.2$; range 3–503).

3.3. Scores and SREs

The mean total score on the START Strength scale was 16.8 ($SD=6.6$; range 4–31); the mean START Vulnerability total score was 24.9 ($SD=7.0$; range 10–40). Scores on the historical scale of the HCR-20 ranged from 5 to 18 (Mean 13.4, $SD=3.2$). The distributions of the SREs are presented in Table 1; in all cases, the low risk category contained the largest proportion of people.

3.4. Base rates

The percentage of people with a lifetime history of each of the adverse outcomes is presented in Table 1; the most common historical behaviour was violence, followed by self-neglect. Base rates for engaging in adverse outcomes in the three months prior to and after START risk assessment are presented in Table 2. The most commonly recorded outcome was any aggression, in which approximately two thirds of the sample engaged in at both time points; the least commonly recorded outcome was self-neglect.

3.5. Predicting SREs from START assessments and recent adverse outcomes

Results of the multinomial logistic regressions are presented in Table 3.

3.5.1. Model for SRE violence

The model examining which components predict violence risk level assignment was significant ($\chi^2[10]=39.74$, $p<.001$); Cox and Snell's $R^2=.40$, Nagelkerke's $R^2=.46$. In detail, the odds of being rated as moderate risk compared to low risk was 0.30 for those with a recorded lifetime history of violence relative to those without a lifetime history; while the vulnerability score was the

Table 1
Risk levels assigned by the specific risk estimates and lifetime history of outcomes.

Risk estimate	Low	Moderate	High	Missing	History
Violence	36 (42.9%)	28 (33.3%)	13 (15.5%)	7 (8.3%)	49 (58.3%)
Self-harm	43 (51.2%)	16 (19%)	5 (6%)	20 (23.8%)	30 (35.7%)
Suicide	42 (50%)	10 (11.9%)	2 (2.4%)	30 (35.7%)	19 (22.6%)
Victimisation	43 (51.2%)	20 (23.8%)	11 (13.1%)	10 (11.9%)	37 (44%)
Self-neglect	41 (48.8%)	26 (31%)	8 (9.5%)	9 (10.7%)	38 (45.2%)

Table 2
Base rates of adverse outcomes.

Outcome	3 Months prior to assessment	3 Months after assessment
Any aggression	54 (64.3%)	57 (67.9%)
Self-harm/suicide	16 (19%)	22 (26.2%)
Victimisation	16 (19%)	18 (21.4%)
Self-neglect	9 (10.7%)	11 (13.1%)

only significant predictor of increased risk of being rated high risk compared with being rated as low risk, and each 1 point increase on this scale increased the odds of being rated as high risk by 1.43. None of the remaining components significantly predicted SRE assignment.

3.5.2. Model for SRE self-harm

The model for prediction of the self-harm SRE was also significant ($\chi^2[10]=21.94$, $p=0.015$); Cox and Snell's $R^2=0.29$, Nagelkerke's $R^2=0.36$. The odds of being rated moderate risk compared to low risk if self-harm/suicidal behaviours were present in the three months prior to the START assessment was 18.35; similarly, the odds of being rated high risk compared to low risk if such behaviours were present in the three month prior to assessment was 20.70. None of the remaining components were significant predictors of the SRE.

3.5.3. Model for SRE suicide

The overall model for predicting risk levels assigned by the suicide SRE was statistically significant ($\chi^2[10]=19.73$, $p=0.032$); Cox and Snell's $R^2=0.31$, Nagelkerke's $R^2=0.43$. Consistent with the results from the self-harm SRE model, the only significant predictor was self-harm/suicidal behaviours in the three months prior to assessment. The odds of being rated moderate risk compared to low risk if recent self-harm/suicidal behaviours were present was 50; this was non-significant when comparing those rated as high with those rated as low risk.

3.5.4. Model for SRE victimisation

The overall model examining which components predict the victimisation SRE was significant ($\chi^2[10]=25.18$, $p=.005$); Cox and Snell's $R^2=0.29$, Nagelkerke's $R^2=0.34$. However, none of the included components significantly predicted assignment to the moderate risk category. For each one-point increase in H10 score, individuals were 1.41 times more likely to be rated as high risk; the odds of those who had been victimised in the three months prior to the START assessment being rated as high risk was 8.24, compared with those who had not been victimised.

3.5.5. Model for SRE self neglect

The model examining components predicting the self-neglect specific risk estimate was non-significant ($\chi^2[10]=18.14$, $p=0.053$); Cox and Snell's $R^2=0.22$, Nagelkerke's $R^2=0.25$.

3.6. Predicting adverse outcomes from START assessment, recent adverse outcomes, and specific risk estimates

The results of the binary logistic regression are presented in Table 4.

3.6.1. Any aggression

Step 1 of the model, based on START components and recent aggressive behaviour, was a significant predictor of any aggression in the three months following assessment ($\chi^2[5]=32.12$, $p<0.001$); Cox and Snell's $R^2=.34$, Nagelkerke's $R^2=0.48$. The overall model after the addition of the specific risk estimate for

Table 3
Multinomial logistic regression results for modelling of specific risk estimates.

	b (SE)	Odds ratio	95%CI
Violence SRE			
[1] Moderate vs. low risk			
Intercept	−2.09 (2.88)		
Strength Score	−0.09 (0.70)	0.91	[0.80, 1.05]
Vulnerability Score	0.05 (0.07)	1.06	[0.92, 1.21]
H10	0.18 (0.10)	1.19	[0.98, 1.46]
History of violence	−1.21 (0.62)*	0.30	[0.09, 1.00]
Aggression in previous 3 months	0.66 (0.65)	1.93	[0.54, 6.89]
[2] High vs. low risk			
Intercept	−13.72 (5.03)**		
Strength Score	0.04 (0.10)	1.04	[0.85, 1.26]
Vulnerability Score	0.36 (0.12)**	1.43	[1.13, 1.80]
H10	0.18 (0.13)	1.20	[0.93, 1.55]
History of violence	−0.83 (0.87)	0.43	[0.08, 2.37]
Aggression in previous 3 months	0.73 (0.99)	2.08	[0.30, 14.34]
Self-harm SRE			
[1] Moderate vs. low risk			
Intercept	1.69 (3.87)		
Strength Score	−0.05 (0.09)	0.95	[0.80, 1.13]
Vulnerability Score	−0.03 (0.09)	0.97	[0.82, 1.15]
H10	−0.16 (0.12)	0.86	[0.67, 1.09]
History of self-harm	0.11 (0.76)	1.12	[0.25, 4.96]
Self-harm/suicide in previous 3 months	2.91 (0.86)***	18.35	[3.37, 99.81]
[2] High vs. low risk			
Intercept	2.83 (6.54)		
Strength Score	−0.14 (0.15)	0.87	[0.64, 1.17]
Vulnerability Score	−0.08 (0.14)	0.93	[0.70, 1.22]
H10	−0.09 (0.20)	0.92	[0.63, 1.35]
History of self-harm	−1.58 (1.47)	0.21	[0.01, 3.66]
Self-harm/suicide in previous 3 months	3.03 (1.36)*	20.70	[1.46, 294.38]
Suicide SRE			
[1] Moderate vs. low risk			
Intercept	−4.43 (5.50)		
Strength Score	0.12 (0.14)	1.13	[0.86, 1.49]
Vulnerability Score	0.01 (0.12)	1.01	[0.80, 1.29]
H10	−0.10 (0.19)	0.91	[0.63, 1.32]
History of suicide ^a	1.20 (1.09)	3.32	[0.40, 27.98]
Self-harm/suicide in previous 3 months	3.91 (1.25)**	50.11	[4.32, 581.35]
[2] High vs. low risk			
Intercept	−3.63 (9.05)		
Strength Score	−0.12 (0.26)	0.89	[0.54, 1.47]
Vulnerability Score	−0.02 (0.22)	0.98	[0.64, 1.51]
H10	0.10 (0.37)	1.11	[0.53, 2.31]
History of suicide ^a	1.48 (1.75)	4.38	[0.14, 134.97]
Self-harm/suicide in previous 3 months	2.11 (2.00)	8.25	[0.16, 418.55]
Victimisation SRE			
[1] Moderate vs. low risk			
Intercept	1.63 (2.87)		
Strength Score	−0.02 (0.07)	0.98	[0.86, 1.12]
Vulnerability Score	−0.01 (0.06)	0.99	[0.88, 1.13]
H10	−0.12 (0.09)	0.89	[0.74, 1.06]
History of victimisation	−0.49 (0.57)	0.61	[0.20, 1.88]
Victimisation in previous 3 months	−0.88 (0.89)	0.42	[0.07, 2.38]
[2] High vs. low risk			
Intercept	−11.87 (5.39)*		
Strength Score	−0.01 (0.12)	0.99	[0.78, 1.26]
Vulnerability Score	0.19 (0.12)	1.21	[0.95, 1.53]
H10	0.35 (0.17)*	1.41	[1.02, 1.95]
History of victimisation	−0.52 (0.91)	0.60	[0.10, 3.56]
Victimisation in previous 3 months	2.11 (0.9)*	8.24	[1.26, 53.96]

Estimates for categorical variables (relevant history and behaviour in previous 3 months) are based on presence of the factor relative to absence of the factor.

^a Defined as suicide, suicide attempts or self-injurious behaviour (Webster et al., 2009).

* $p<.05$.

** $p<.01$.

*** $p<.001$.

violence was also significant ($\chi^2[7]=32.88$, $p < 0.001$); Cox and Snell's $R^2=0.35$, Nagelkerke's $R^2=0.49$; however, the addition of the SRE to the model did not improve prediction ($\chi^2[2]=0.76$, $p=0.683$). Any aggression occurring in the three months before START assessment was the only significant predictor in the model; individuals who had been aggressive were 7.81 times more likely to also be aggressive in the subsequent three months compared with those who had not been aggressive prior to the assessment. This was reduced to 7.50 times more likely after the addition of the SRE in step 2. The violence SRE did not predict its corresponding outcome.

3.6.2. Self-harm/suicide

Step 1, based on START components and recent self-harm and suicidal behaviour, was a significant predictor of self-harm/suicide in the three months following assessment ($\chi^2[6]=20.70$, $p=0.002$); Cox and Snell's $R^2=.32$, Nagelkerke's $R^2=0.45$. The overall model after the addition of the specific risk estimates for self-harm and suicide was also significant ($\chi^2[10]=31.92$, $p < 0.001$); Cox and Snell's $R^2=0.45$, Nagelkerke's $R^2=0.64$ and the addition of the SREs added incremental predictive utility ($\chi^2[4]=11.23$, $p=0.024$). Again, the only significant predictor was self-harm/suicide in the preceding three months, which increased odds of engaging in subsequent self-harm/suicide by 22.26; this was non-significant after the addition of the SREs in step 2. Neither the self-harm or suicide SREs were significant predictors of the composite self-harm/suicide outcome.

3.6.3. Victimisation

Step 1, based on START components and recent victimisation, was a significant predictor of victimisation in the three months following assessment ($\chi^2[5]=16.80$, $p=0.005$); Cox and Snell's $R^2=.20$, Nagelkerke's $R^2=0.31$. The overall model after the addition of the SRE for victimisation was also significant ($\chi^2[7]=30.20$, $p < 0.001$); Cox and Snell's $R^2=0.34$, Nagelkerke's $R^2=0.51$ and the addition of the SRE added incremental predictive utility ($\chi^2[2]=13.40$, $p=0.001$). The only significant predictor was victimisation in the three months before assessment which increased odds of further victimisation by 11.73; this was reduced to 8.69 times more likely after the addition of the SRE. The odds of being victimised following the START assessment was 34.48 times more likely for those rated high risk relative to those rated low risk.

3.6.4. Self-neglect

Step 1, based on START components and recent self-neglect, was a significant predictor of self-neglect in the three months following assessment ($\chi^2[5]=22.26$, $p < 0.001$); Cox and Snell's $R^2=.26$, Nagelkerke's $R^2=0.45$. The overall model after the addition of the specific risk estimate for self-neglect was also significant ($\chi^2[7]=23.68$, $p < 0.001$); Cox and Snell's $R^2=0.27$, Nagelkerke's $R^2=0.48$. The addition of the SRE did not add incremental predictive utility ($\chi^2[2]=1.42$, $p=0.491$). Self-neglect in the three months before assessment was a significant predictor of self-neglect in the three months after START assessment and increased odds of engaging by 23.58; this was increased to 25.08 after the addition of the SRE in step 2. The Strength score was also a significant predictor; for every 1 point increase on this scale (i.e. as the number of strengths increased) individuals were 0.70 times as likely to engage in self-neglect. The SRE was not a significant predictor of its corresponding outcome.

4. Discussion

We have investigated how clinicians in a secure psychiatric setting make formulations of risk, and how those formulations fair

Table 4

Binary logistic regression results for predicting adverse outcomes based on START components and specific risk estimates.

	b (SE)	Odds ratio	95%CI
Any aggression			
Step 1			
Strength score	−0.14 (0.08)	0.87	[0.74, 1.01]
Vulnerability score	−0.02 (0.07)	1.00	[0.87, 1.15]
H10	0.21 (0.12)	1.24	[0.99, 1.55]
History of violence	−1.15 (0.76)	0.32	[0.07, 1.41]
Aggression in previous 3 months	2.06 (0.65)**	7.81	[2.19, 27.90]
Constant	0.16 (3.14)	1.18	
Step 2			
Strength score	−0.14 (0.08)	0.87	[0.75, 1.02]
Vulnerability score	−0.01 (0.08)	0.99	[0.85, 1.15]
H10	0.19 (0.12)	1.21	[0.96, 1.53]
History of violence	−1.07 (0.78)	0.34	[0.07, 1.60]
Aggression in previous 3 months	2.02 (0.66)**	7.50	[2.08, 27.08]
Violence SRE [1]	0.67 (0.78)	1.96	[0.43, 8.96]
Violence SRE [2]	0.31 (1.10)	1.36	[0.16, 11.71]
Constant	0.30 (3.20)	1.34	
Self-harm/suicide			
Step 1			
Strength score	−0.12 (0.11)	0.89	[0.72, 1.11]
Vulnerability score	−0.12 (0.10)	0.88	[0.72, 1.08]
H10	0.19 (0.16)	1.21	[0.88, 1.66]
History of self-harm	−1.01 (1.41)	0.37	[0.02, 5.79]
History of suicide ^a	1.02 (1.41)	2.78	[0.18, 43.71]
Self-harm/suicide in previous 3 months	3.10 (1.00)**	22.26	[3.13, 158.29]
Constant	0.61 (4.47)	1.84	
Step 2			
Strength score	−0.20 (0.15)	0.82	[0.61, 1.10]
Vulnerability score	−0.15 (0.13)	0.86	[0.67, 1.12]
H10	0.28 (0.18)	1.32	[0.92, 1.89]
History of self-harm	0.21 (1.66)	1.23	[0.05, 31.61]
History of suicide ^a	−1.23 (1.90)	0.29	[0.01, 12.04]
Self-harm/suicide in previous 3 months	1.73 (1.27)	5.65	[0.47, 67.61]
Self-harm SRE [1]	0.33 (1.65)	1.39	[0.05, 35.31]
Self-harm SRE [2]	0.37 (3.36)	1.44	[0.00, 1039.06]
Suicide SRE [1]	3.16 (1.78)	23.52	[0.72, 772.12]
Suicide SRE [2] ^b			
Constant	0.85 (5.34)	2.33	
Victimisation			
Step 1			
Strength score	−0.02 (0.08)	0.98	[0.83, 1.15]
Vulnerability score	0.06 (0.08)	1.06	[0.91, 1.23]
H10	0.13 (0.11)	1.14	[0.92, 1.42]
History of victimisation	0.13 (0.67)	1.13	[0.31, 4.18]
Victimisation in previous 3 months	2.46 (0.72)***	11.73	[2.87, 47.88]
Constant	−4.81 (3.26)	0.01	
Step 2			
Strength score	−0.03 (0.09)	0.97	[0.81, 1.16]
Vulnerability score	−0.02 (0.09)	0.98	[0.83, 1.16]
H10	−0.02 (0.14)	0.98	[0.75, 1.28]
History of victimisation	0.31 (0.78)	1.36	[0.30, 6.25]
Victimisation in previous 3 months	2.16 (0.82)**	8.69	[1.73, 43.75]
Victimisation SRE [1]	−0.16 (0.97)	0.85	[0.13, 5.64]
Victimisation SRE [2]	3.54 (1.16)**	34.48	[3.58, 332.27]
Constant	−1.37 (3.69)	0.26	
Self-neglect			
Step 1			
Strength score	−0.38 (0.15) [†]	0.69	[0.52, 0.92]

Table 4 (continued)

	b (SE)	Odds ratio	95%CI
Vulnerability score	−0.18 (0.10)	0.84	[0.68, 1.03]
H10	0.22 (0.19)	1.25	[0.86, 1.81]
History of self-neglect	−1.14 (0.90)	0.32	[0.06, 1.88]
Self-neglect in previous 3 months	3.16 (1.13)**	23.58	[2.59, 215.11]
Constant	4.98 (3.90)	144.74	
Step 2			
Strength score	−0.36 (0.15)*	0.70	[0.52, 0.93]
Vulnerability score	−0.19 (0.11)	0.83	[0.67, 1.02]
H10	0.22 (0.19)	1.25	[0.86, 1.82]
History of self-neglect	−1.05 (0.91)	0.35	[0.06, 2.11]
Self-neglect in previous 3 months	3.22 (1.15)**	25.08	[2.63, 238.95]
Self-neglect SRE [1]	1.03 (0.97)	2.81	[0.42, 18.97]
Self-neglect SRE [2]	1.15 (1.29)	3.16	[0.25, 39.57]
Constant	4.34 (3.93)	76.43	

Estimates for categorical variables (relevant history and behaviour in previous 3 months) are based on presence of the factor relative to absence of the factor.

SRE [1]=moderate risk relative to low risk.

SRE [2]=high risk relative to low risk.

^a Defined as suicide, suicide attempts or self-injurious behaviour (Webster et al., 2009).

^b Unable to calculate parameter estimates.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

as predictors of adverse outcomes compared with their constituent parts. Rating teams considered historical violent behaviour to be important components of an elevated SRE rating of moderate. The presence of dynamic risk factors, represented by the START Vulnerabilities scale, was important in the decision to make an SRE rating of high risk but did not play a significant role in a moderate rating. However, the only component in the START risk formulation to significantly predict actual aggression in the three months following assessment was the presence of aggression in the three month period prior to assessment. As a result, we can conclude that for this outcome, the clinical teams did not consider the available information in the optimal manner since their rating of increased risk would have most accurately predicted outcome if it had been based solely on recent behaviour. This finding is consistent with previous research about the START. First, clinicians are considering dynamic risk factors in addition to a history of violence when formulating the violence SRE; this is probably because they know that this is where the greatest evidence base lies (e.g., O'Shea and Dickens, 2014). Second, they have some difficulty in applying the information to the case at hand specifically in making fine-grained decisions about the level of risk (Doyle et al., 2008) since the dynamic factors only appear to contribute to assignment to the high risk category. Too much emphasis is being placed on historical factors; the current evidence suggests that patients with a lifetime history of violence will be unlikely to be rated at low risk of violence irrespective of recent behaviour and scores on dynamic factors. These factors have extremely limited ability to demonstrate positive change as a result of therapeutic interventions (Rogers, 2000), such that patients likely score highly in these domains following treatment, irrespective of reduction in dynamic risk factors and aggressive incidents. The current results therefore suggest that if a patient is rated at elevated risk of violence when admitted to a secure psychiatric setting due to history of violence, this rating will be unlikely to decrease. Indeed, this hypothesis has been borne out by recent HCR-20 longitudinal research (Morrissey et al., 2014;

O'Shea and Dickens, 2015a). Further, results also suggest that a START SRE of high risk for violence in the absence of recent violence may be unwarranted given the relative predictive value of the latter; as a result it is possible that restrictive management practices based on such an estimate is not attuned to the individual's actual risk and could therefore be unethical.

For the self-harm and suicide SREs, the only component that significantly predicted the assigned risk level was the presence of a corresponding adverse event in the preceding three months suggesting that clinicians are giving the most weight to previous behaviour and less to risk and protective factors. This is consistent with previous research which has shown that START SREs, but not dynamic factors, predict their associated outcomes (O'Shea and Dickens, 2014; O'Shea et al., 2016). Clinicians should be advised to maintain focus on recent corresponding adverse events when formulating SREs for self-harm and suicide as, for these outcomes, the only component to significantly predict each outcome was the presence of the corresponding behaviour in the three months before assessment. In the case of victimisation recent adverse outcome was a significant predictor of outcome although the SRE for victimisation added predictive value. These results again raise doubt regarding the relevance of the START items and the H10 items for the prediction of self-harm, suicide and victimisation. In fact, the START authors detail the rationale for including each of the items, but rarely cite empirical links between the items and outcomes other than aggression or violence (Webster et al., 2009). Whilst reliance on recent behaviour may be useful for identifying those who are at risk, it does not provide useful treatment targets and ultimately the aim of risk assessment is to inform prevention strategies (Douglas and Kropp, 2002). The recreational and impulse control items in the START have been found to be predictive of self-harm/suicide in males, whereas substance use, rule adherence and coping are of relevance to females (O'Shea and Dickens, 2015b). However, further work is needed to determine if any of the individual START items are of relevance to victimisation, and to derive additional evidence based risk and protective factors that can be incorporated into a structured risk assessment.

The self-neglect SRE was also only significantly predicted by the presence of self-neglect in the three months prior to START assessment. However, the Strength score was a significant predictor of self-neglect following assessment, in addition to recent behaviour, but the SRE did not predict its associated outcome. This is consistent with recent findings that demonstrate a lack of predictive efficacy of the SRE despite significant prediction achieved by the START Strength scores (O'Shea et al., 2016), and suggests that Strength scores may be of particular relevance to self-neglect, but are not considered as such by clinicians. Whilst self-neglect may be less problematic in psychiatric inpatients, in comparison with those in the community, as basic needs are provided for (Gunstone, 2003), relatively high rates have been reported among hospitalised patients (Lelliott et al., 1994). It would be beneficial to determine whether interventions aimed at bolstering protective factors are useful in reducing self-neglect in secure settings and whether advising clinicians to consider protective factors when formulating their SRE for self-neglect improves its predictive accuracy.

4.1. Limitations

Risk assessments were completed as part of routine clinical practice; therefore, we were unable to obtain any qualitative information from mental health professionals regarding the information that they believe is most important in influencing their SRE. It would be beneficial to obtain such information to support the current results as examining which components are statistically related to the specific risk estimate only acts as a proxy for

what clinicians are actually considering. Further, these findings come from an inpatient setting where information about recent behaviour is readily accessible and there is a strong culture of recording of adverse events. Therefore, these findings may not be generalizable to decision-making in other settings, where clinicians may not have access to recent behaviours, and hence, place a greater reliance on other components. Due to the restricted sample size we were only able to evaluate the contribution of a limited number of structured judgement components and, as a result, we are not able to comment on whether and how other component items such as key strengths, critical vulnerabilities, or T.H.R.E.A.T. items contribute to assessment. It is not entirely clear, however, how they might be operationalized as variables. Given the very large number of variables suggested for consideration in rating the START, the important role of the individual clinician in the process, and the requirement to consider case-specific factors we question whether it will ever be possible to adequately investigate the precise contribution of all items to risk estimation. In the current study setting, the START and HCR-20 are mandated for all adult patients; therefore, the current sample should be largely representative of the wider patient population. However, as patients were required to be resident in the service for 3 months before and after START assessment to allow for the collection of outcome data, the included sample may not be representative of patients whose admission was shorter than 3 months, and the study was conducted across one provider of mental health services. Therefore, practitioners should consider the sample characteristics when interpreting the generalisability of results to their own practice, particularly for short-stay patients. Finally, the amount of missing SREs ranged from 8.3 (SRE violence) to 35.7% (SRE suicide) and it is possible that this might affect generalisability.

4.2. Conclusions

The current study provides some of the first insights into the formulation of SREs in a secure psychiatric setting and suggests that clinicians are not considering available information in the optimal manner for all outcomes. In total, the results suggest very limited utility of the START components as predictors of adverse outcomes beyond what is achieved by recent behaviours. While the examination of the predictive efficacy of assessment tools like the START is a necessary step in determining their ability to assist in the prevention and management of adverse outcomes, the future of research with the START should be in determining if it provides a useful structure to allow teams to consider possible intervention targets, and ultimately reduce the number of negative outcomes. For outcomes like self-harm and victimisation the majority of the current START items are unlikely to be informative. For the victimisation outcome, SREs made by teams seems to incorporate factors that provide some incremental validity over predictions based solely on recent behaviour; we cannot distinguish what these factors are from the current design and this should be subject to further research.

References

- Buchanan, A., 2013. Violence risk assessment in clinical settings: being sure about being sure. *Behav. Sci. Law* 31 (1), 74–80. <http://dx.doi.org/10.1002/bsl.2045>.
- Crocker, A.G., Braithwaite, E., Laferriere, D., Gagnon, D., Venegas, C., Jenkins, T.A., 2011. START changing practice: implementing a risk assessment and management tool in a civil psychiatric setting. *Int. J. Forensic Ment. Health* 10 (1), 13–28. <http://dx.doi.org/10.1080/14999013.2011.553146>.
- de Ruiter, C., Nicholls, T.L., 2011. Protective factors in forensic mental health: a new frontier. *Int. J. Forensic Ment. Health* 10 (3), 160–170. <http://dx.doi.org/10.1080/14999013.2011.600602>.
- Douglas, K., Hart, S., Webster, C., Belfrage, H., 2013. HCR-20: Assessing Risk for Violence (Version 3). Mental Health, Law, and Policy Institute, Simon Fraser University, Burnaby, BC.
- Douglas, K.S., Kropp, P.R., 2002. A prevention-based paradigm for violence risk assessment – clinical and research applications. *Crim. Justice Behav.* 29 (5), 617–658. <http://dx.doi.org/10.1177/009385402236735>.
- Doyle, M., Lewis, G., Brisbane, M., 2008. Implementing the Short-Term Assessment of Risk and Treatability (START) in a forensic mental health service. *Psychiatr. Bull.* 32 (11), 406–408. <http://dx.doi.org/10.1192/pb.bp.108.019794>.
- Dvoskin, J.A., Heilbrun, K., 2001. Risk assessment and release decision-making: Toward resolving the great debate. *J. Am. Acad. Psychiatry Law* 29 (1), 6–10.
- Eberhart, N.K., Hammen, C.L., 2006. Interpersonal predictors of onset of depression during the transition to adulthood. *Pers. Relatsh.* 13 (2), 195–206. <http://dx.doi.org/10.1111/j.1475-6811.2006.00113.x>.
- Elbogen, E.B., 2002. The process of violence risk assessment: a review of descriptive research. *Aggress. Violent Behav.* 7 (6), 591–604. [http://dx.doi.org/10.1016/S1359-1789\(01\)00051-9](http://dx.doi.org/10.1016/S1359-1789(01)00051-9).
- Gunstone, S., 2003. Risk assessment and management of patients whom self-neglect: a 'grey area' for mental health workers. *J. Psychiatr. Ment. Health Nurs.* 10 (3), 287–296. <http://dx.doi.org/10.1046/j.1365-2850.2003.00568.x>.
- Guy, L.S., Packer, I.K., Warnken, W., 2012. Assessing risk of violence using structured professional judgment guidelines. *J. Forensic Psychol. Pract.* 12 (3), 270–283. <http://dx.doi.org/10.1080/15228932.2012.674471>.
- Hart, S.D., Logan, C., 2011. Formulation of violence risk using evidence-based assessments: the structured professional judgment approach. In: Sturmey, P., McMurran, M. (Eds.), *Forensic Case Formulation*. Wiley-Blackwell, Chichester, UK.
- Large, M.M., Ryan, C.J., Singh, S.P., Paton, M.B., Nielssen, O.B., 2011. The predictive value of risk categorization in schizophrenia. *Harv. Rev. Psychiatry* 19 (1), 25–33. <http://dx.doi.org/10.3109/10673229.2011.549770>.
- Lelliott, P., Wing, J., Clifford, P., 1994. A national audit of new long-stay psychiatric patients. I: Method and description of the cohort. *Br. J. Psychiatry* 165 (2), 160–169. <http://dx.doi.org/10.1192/bjp.165.2.160>.
- Morrissey, C., Beeley, C., Milton, J., 2014. Longitudinal HCR-20 scores in a high-secure psychiatric hospital. *Criminal. Behav. Ment. Health* 24 (3), 169–180. <http://dx.doi.org/10.1002/cbm.1893>.
- Nagelkerke, N.J., 1991. A note on a general definition of the coefficient of determination. *Biometrika* 78 (3), 691–692.
- National Institute for Mental Health in England, 2004. Mental Health Policy Implementation Guide. Retrieved 1st March 2012 from (http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4087136).
- Nicholls, T.L., Gagnon, N., Crocker, A.G., Brink, J., Desmarais, S.L., Webster, C., 2007. START Outcomes Scale (SOS). BC Mental Health and Addiction Services, Vancouver.
- O'Shea, L.E., Dickens, G.L., 2014. Short-Term Assessment of Risk and Treatability (START): systematic review and meta-analysis. *Psychol. Assess.* 26 (3), 990–1002. <http://dx.doi.org/10.1037/a0036794>.
- O'Shea, L.E., Dickens, G.L., 2015a. The HCR-20 as a measure of reliable and clinically significant change in violence risk among secure psychiatric inpatients. *Compr. Psychiatry* 62, 132–140. <http://dx.doi.org/10.1016/j.comppsy.2015.07.009>.
- O'Shea, L.E., Dickens, G.L., 2015b. Predictive validity of the Short-Term Assessment of Risk and Treatability (START) for aggression and self-harm in a secure mental health service: gender differences. *Int. J. Forensic Ment. Health* 14 (2), 132–146. <http://dx.doi.org/10.1080/14999013.2015.1033112>.
- O'Shea, L.E., Piccioni, M.M., Dickens, G.L., 2016. The predictive validity of the Short-Term Assessment of Risk and Treatability (START) for multiple adverse outcomes in a secure psychiatric inpatient setting. *Assessment* 23, 150–162. pii: 1073191115573301.
- Oullette, J.A., Wood, W., 1998. Habit and intention in everyday life: the multiple processes by which past behavior predicts future behavior. *Psychol. Bull.* 124 (1), 54–74.
- Peduzzi, P., Concato, J., Kemper, E., Holford, T.R., Feinstein, A.R., 1996. A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* 49 (12), 1372–1379.
- Peng, C.-Y.J., Lee, K.L., Ingersoll, G.M., 2002. An introduction to logistic regression analysis and reporting. *J. Educ. Res.* 96 (1), 3–14.
- Rogers, R., 2000. The uncritical acceptance of risk assessment in forensic practice. *Law Hum. Behav.* 24 (5), 595–605.
- Skeem, J., Schubert, C., Stowman, S., Beeson, S., Mulvey, E., Gardner, W., Lidz, C., 2005. Gender and risk assessment accuracy: Underestimating women's violence potential. *Law Hum. Behav.* 29 (2), 173–186.
- Szmukler, G., Rose, N., 2013. Risk assessment in mental health care: values and costs. *Behav. Sci. Law* 31 (1), 125–140. <http://dx.doi.org/10.1002/bsl.2046>.
- Webster, C.D., Douglas, K.S., Eaves, D., Hart, S.D., 1997. HCR-20: Assessing Risk for Violence (version 2). Mental Health, Law and Policy Institute, Simon Fraser University, Burnaby, British Columbia.
- Webster, C.D., Martin, M.L., Brink, J., Nicholls, T.L., Desmarais, S.L., 2009. Manual for the Short-Term Assessment of Risk and Treatability (START) (Version 1.1). British Columbia Mental Health and Addiction Services, Coquitlam, Canada.
- World Health Organisation, 1992. The ICD-10 Classification of Mental and Behavioural Disorders. WHO, Geneva.