

How short should short-term risk assessment be? Determining the optimum interval for START reassessment in a secure mental health service

Geoffrey L. Dickens and Laura E. O'Shea

This is the peer reviewed version of the following article:

Dickens, G.L. and O'Shea, L.E. 2015. How short should short-term risk assessment be? Determining the optimum interval for START reassessment in a secure mental health service. *Journal of Psychiatric and Mental Health Nursing*. 22(6): pp.397-406. doi: 10.1111/jpm.12232

which has been published in final form at:

<http://dx.doi.org/10.1111/jpm.12232>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

How short should short-term risk assessment be? Determining the optimum interval for START reassessment in a secure mental health service

Geoffrey L. Dickens* RMN BSc(Hons) PGDip MA PhD

St Andrew's Academic Department, St Andrew's Hospital, Northampton. NN1 5DG. United Kingdom and School of Health, University of Northampton, Northampton. NN2 7AL. United Kingdom. gdickens@standrew.co.uk

Laura E. O'Shea BSc(Hons)

St Andrew's Academic Department, St Andrew's Hospital, Northampton. NN1 5DG. United Kingdom and King's College London Institute of Psychiatry, 16 De Crespigny Park, London, SE5 8AF. United Kingdom. leoshea@standrew.co.uk

*Corresponding author

ABSTRACT

Aim: The Short Term Assessment of Risk and Treatability (START) guides assessment of potential adverse outcomes. Assessment is recommended every 3-months but there is no evidence for this interval. We aimed to inform whether earlier reassessment was warranted.

Method: We collated START assessments for $N=217$ adults in a secure mental health hospital, and subsequent aggressive, self harm, self neglect and victimisation incidents. We used Receiver Operating Characteristic analysis to assess predictive validity; survival function analysis to examine differences between low-, medium-, and high-risk groups; and hazard function analysis to determine the optimum interval for reassessment.

Results: The START predicted aggression and self-harm at 1-, 2- and 3- months. At-risk individuals engaged in adverse outcomes earlier than low risk patients. About half warranted reassessment before 3-months due to engagement in risk behaviour before that point despite a low risk rating; or because of non-engagement by that point despite an elevated risk rating.

Discussion: Risk assessment should occur at appropriate intervals so that management strategies can be individually tailored.

Implications for practice: Assessment at 3-month intervals is supported by the evidence. START assessments should be revisited earlier if risk behaviours are not prevented; teams should constantly re-evaluate the need for restrictive practices.

Keywords: Risk assessment, risk management, aggression, violence, self harm

ACCESSIBLE SUMMARY

What is known on the subject

- The Short Term Assessment of Risk and Treatability (START) is a tool used in some mental health services to assess patients to see if they are at risk of violence, self-harm, self-neglect, or victimisation.
- The recommended time between assessments is 3-months but there is currently no evidence to show that this is best practice.

What this paper adds to existing knowledge

- We have investigated whether assessing at 1- or 2- month intervals would be more accurate and therefore facilitate more individualised risk management interventions.
- We found that many patients who were rated as low risk had been involved in risk behaviours before 3-months had passed; some patients who were rated at increased risk did not get involved in risk behaviours at all.
- This study provides the first international evidence about the appropriate time for START reassessment.

What are the implications for practice

- Results are mixed for different outcomes but on balance we think that the recommendation to conduct START assessment every 3-months is supported by the evidence. However, reassessment should be considered if risk behaviours are not prevented and teams should always consider whether risk management practices are too restrictive.

.INTRODUCTION

Risk management is a key role of mental health nurses working in secure/forensic inpatient services. The Short Term Assessment of Risk and Treatability (START; Webster et al., 2009) is a structured clinical judgement scheme which aims to guide the assessment of patients with mental illness and/or personality disorder for a range of adverse outcomes through the consideration of dynamic risk and protective factors (termed, respectively, Vulnerabilities and Strengths). The tool was developed in Canada but has been used clinically and researched in the United Kingdom, Australia, Ireland and Norway (O'Shea & Dickens, 2014). The START is feasible to complete and has utility and value in the risk management process (Crocker et al., 2008, 2011; Desmarais et al., 2011; Doyle et al., 2008; Quinn et al., 2013; van den Brink et al., 2009). Inter-rater reliability of the Strength and Vulnerability scales and the specific risk estimates is good to excellent (mean intraclass correlations .78-.86), and the tool has convergent validity with other instruments that aim to assess risk for aggression and self-harm (O'Shea & Dickens, 2014). It has the potential to support recovery-focused forensic mental health nursing care (Dickens & Doyle, in press; Doyle & Jones, 2013). Evidence for the predictive validity of the START is mixed: the overall specific risk estimates (SREs), which are the authors' recommended method of judging overall risk, are good predictors of aggressive and self-harm outcomes; while the Strengths and Vulnerabilities scales are good predictors of aggressive and violent outcomes but not of self-harm (O'Shea & Dickens, 2014).

START is intended to predict outcomes over a period of up to 3 months; this is in contrast to, for example, both the HCR-20 (Webster, Douglas, Eaves & Hart, 1997) and the Risk for Sexual Violence Protocol (Hart et al., 2003) which are to be repeated at 6-12 month intervals;

and ultra-short term assessments of violence risk such as the Dynamic Appraisal of Situational Aggression (DASA; Ogloff & Daffern, 2006) which assesses risk for the subsequent 24-hour period. No rationale is offered in the START manual (Webster et al., 2009) for the selection of a 3-month inter-assessment period. Studies of the predictive validity of the START have used follow-up periods lasting between 30 days and 1 year (O'Shea & Dickens, 2014), but there is currently insufficient data to determine via meta-analysis whether length of follow-up period moderates the effect size of predictive outcomes. Previous studies have partially addressed this question. Chu et al. (2011) found the START Vulnerabilities scale to be a significant predictor of aggressive outcomes at periods of 1-, 3-, and 6-months (Area Under the Curve [AUC] range .74 - .83) but did not examine the tool's Strengths scale, its SREs, or non-aggressive outcomes. Wilson et al. (2011, 2013) have reported statistically significant AUC values for the Strengths scale, Vulnerabilities scale, and SRE for violence at 3-, 6-, 9- and 12- month periods following assessment; again, these studies only examined aggressive and sexually inappropriate behaviours and not the full range of START outcomes. This is an important question since more accurate prediction over one period of time compared with another would have implications for the frequency with which reassessment is conducted. An answer would have clinical value since risk management procedures that involve restrictive practices should be calibrated to the individual's level of risk at any given time. Further, reassessment at too short intervals may needlessly drain valuable resources. It is also important to consider the period following assessment in which patients remain event-free in terms of risk outcomes (also known as the survival period) since such engagement might of itself warrant reassessment.

While the START has been used internationally, and there is some evidence for its feasibility, utility and predictive value, there is currently no evidence about the appropriate interval between assessments. The current study aimed to determine whether reassessment at an earlier point than that recommended in the START manual is warranted. The specific objectives were, first, to examine the predictive validity of the START at 1-, 2- and 3- month temporal gates following assessment; second, to examine survival function by assigned risk level for the various outcomes across the same temporal gates; finally, to perform hazard function analysis to determine the optimum point for reassessment of each outcome.

METHOD

Participants

The study was conducted at St. Andrew's, a UK-based, independent sector provider of secure inpatient mental health care at four sites. Accommodation is provided in levels of low and medium security for people diagnosed with mental disorder, intellectual disability, and acquired brain injury. All patients in the current study were previously reported on by O'Shea, Picchioni, and Dickens (2014). Eligible patients were adults (aged 18 years +) consecutive admissions between May 2011 and July 2012 who had a START risk assessment completed by their clinical team. Patients were excluded if their START assessment had missing data in excess of guidance in the START manual (Webster et al., 2009).

Procedure

We used a pseudo-prospective study design; START assessments were completed by multidisciplinary teams as part of routine clinical practice and risk incidents were recorded in electronic progress notes by qualified clinical staff every shift. Patients' demographic and

clinical characteristics, first START assessment, and risk incidents for the subsequent three months were extracted from electronic records. All data were anonymised and linked by code numbers. The study did not require research ethics approval as anonymised data only were accessed. The study was approved by the Hospital's Clinical Audit Lead.

Measures

START assessment. START assessment was conducted in the course of routine clinical practice according to recommendations in the START manual and collated for the study. The START was designed to be completed by a “number of mental health specialists who work together as a team” (Webster, et al., 2009; p. 24). It comprises 20 items (see Table 1); each is rated both in terms of Vulnerabilities and Strengths on a 3-point scale (0 = no/minimal strength/vulnerability, 1 = moderate strength/vulnerability, 2 = high strength/vulnerability). Raters are advised to indicate key and critical items to identify strengths and vulnerabilities that seem especially important for the case at hand, and to record whether the patient has a history of behaviours relating to each of the seven risk areas that the START aims to address: violence to others, self-harm, suicide, substance abuse, victimisation, self-neglect, and unauthorised absence. SREs (low, moderate, or high risk) regarding the likelihood of each of these outcomes occurring over the subsequent three months are then made. There are no explicit guidelines about how SREs are to be arrived at and it is not the intention of the tool's authors to provide cut-off scores to assist raters; rather, raters are advised to use their clinical judgement to interpret the START assessment information and translate it into a rating of low, moderate or high risk. For research purposes, the total Strength and Vulnerability scales can be summed and, if necessary, prorated to account for missing items following guidelines in the START manual (Webster, et al., 2009).

Raters in the current study setting attended structured theoretical and practical training in START completion involving team discussion and rating of pseudonymised cases. In practice, completed START assessments are signed off by three members of the multidisciplinary team from different professions. These assessments are completed every three months for each patient and are routinely audited to ensure compliance.

Demographic and clinical data. Patients' age, gender, self-reported ethnicity, admission/discharge date, security level, legal status, and ICD-10 (World Health Organisation, 1992) psychiatric diagnoses were extracted from clinical records.

Risk outcomes. For each patient, an electronic progress note is entered on every nursing shift by a qualified member of clinical staff. Notes are flagged if any of a range of risk outcomes occur. As part of a previous study (O'Shea, Picchioni, & Dickens, 2014), incidents flagged for "Aggression – Physical", "Aggression – Verbal", "Self-harm/Suicide", "Self-neglect", "Vulnerability", "Unauthorized leave", and "Substance abuse" were collated. Collated notes were then coded by the authors who were blind to the START assessment using the START Outcome Scale (SOS; T. L. 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*): verbal aggression, aggression against property, physical aggression against others, sexual aggression, self-harm, suicide ideation and planning, suicide behaviors, self-neglect, unauthorized leave, substance abuse, being victimized, and stalking. We did not code entries separately for sexual aggression and stalking since the START is not intended to predict these outcomes. Physical aggression against property and against others was combined into a physical aggression category, and similarly, self-harm, suicide ideation and planning, and suicide behaviours were combined into a self-harm

category. For the current study we counted an outcome-type as 'present' if there was at least one event rated as scoring 1 or above on the SOS in each 1-month outcome period. Inter-rater reliability was in the excellent range: Kappa ranged from .83 to 1.00, the lowest being for self-neglect and the highest for self-harm and physical aggression. We treated aggression against property and physical aggression against others as a single outcome (physical aggression).

Data analysis

Power calculations were conducted in MedCalc for Windows, version 14.8.1 (MedCalc software, Ostend, Belgium) based on an expected large effect size, with $\alpha = .05$ and $\beta = .20$, to identify the required sample size to have sufficient power to detect predictive ability of the START for each outcome. Power calculations revealed that 824 and 501 cases would be required to detect a significant effect for substance abuse and unauthorized leave, respectively, due to low base rates. Since we had access to a maximum of $N=231$ case records the present study was underpowered to detect predictive efficacy for these two outcomes and no further analysis was conducted. There was sufficient power to detect predictive ability for the remaining outcomes; which required sample sizes ranging from 54 for physical aggression to 153 for self-neglect. Descriptive statistics (frequencies, Mean, SD) were used to examine distribution of START scores, SREs, and the occurrence of risk outcomes at three temporal gates (month 1, month 2 and month 3 following assessment). The predictive validity of the START Strengths and Vulnerability scales, SREs, and all individual Strength and Vulnerability items at each temporal gate was examined using Receiver Operating Characteristic analysis (ROC) principles and calculating the Area Under the Curve (AUC) based on sensitivity and specificity. The AUC value ranges from 0 to 1, with .5 representing chance performance while, typically, .75 is

considered the threshold for a large effect size (Dolan & Doyle, 2000). Second, we conducted Kaplan-Meier survival function analysis to determine whether patients designated to low, medium and high SREs differed significantly in terms of their survival (i.e., abstention from engagement in risk outcomes termed). The Log Rank (Mantel Cox) statistic was used to determine the probability of survival curves differing by chance. Mean and 95% Confidence Intervals (CIs) survival times were calculated for each SRE group and compared pairwise to determine significant between-group survival rates. Data for patients who were discharged before gate 1 and gate 2 were censored. Finally, we determined the Hazard function $h(t)$. This quantifies the likelihood of failure to survive at any given time point (t) after assessment for patients who have survived up to that point. It is calculated as the number of patients failing to abstain from risk engagement at any given time point as a proportion of the total population of abstaining patients at the beginning of that time point (see Fleischman et al., 2010). The earliest point at which $h(t)$ fails to decrease may be considered the optimum time for reassessment since little additional relative risk is to be expected at this point. All analyses were conducted using PASW Statistics version 18 for Windows (SPSS Inc., Version 18).

RESULTS

In total, 231 patients met the inclusion criteria; 14 were excluded due to missing data leaving a final sample size of $N=217$ (Mean age = 34.1 years [$SD=15.0$]; 75.1% male; response rate 93.9%). The majority of patients ($n=130$, 59.9%) were resident in low secure wards and 87 (40.1%) were in medium secure wards. Ninety seven patients (44.7%) were detained under forensic sections of the Mental Health Act (1983, 2007), 108 (50.8%) were held under civil sections and 12 (5.5%) were admitted informally. Mean time between admission and START

assessment was 179 days ($SD = 147.7$). The number of patients assigned to each risk level, and the number of patients engaging in each outcome for the first time at each temporal gate, is presented in Table 2. The most common adverse behaviour was verbal aggression followed by physical aggression. AUC values for each outcome-predictor pairing at each temporal gate are detailed in Table 3. Results largely replicated those previously reported for 3-month outcomes only (O'Shea *et al.*, 2015). Strengths and Vulnerabilities total scores and the Aggression SRE were significantly predictive of aggressive physical and verbal aggression at all three temporal gates. Only the self-harm SRE, not the Strengths or Vulnerabilities scores, was predictive of self-harm behaviour. Additionally the SRE for suicide predicted self-harm at all temporal gates. The SRE for victimisation predicted its associated outcome at gate 3 but not at the preceding gates; while the Strength score predicted self-neglect at gates 1 and 2 but not at gate 3. Of 17 statistically significant AUC values at gates 1 and 2 six decreased at the subsequent gate, ten increased, and one was unchanged.

Survival analysis was conducted to examine differences between those designated low, moderate and high risk (see Figures 1 - 4) for all outcomes where SRE was a significant predictor of outcomes; as a result, self-neglect was omitted from this analysis since the AUC values suggested there was no predictive utility for the associated SRE. Survival rates for physical aggression for patients differed significantly between those designated low risk and moderate risk ($\chi^2 = 11.040$, $df=1$, $p=.001$), between those rated low risk and those high risk ($\chi^2 = 25.480$, $df=1$, $p<.001$), and between those rated moderate risk and those high risk ($\chi^2 = 4.274$, $df=1$, $p<.05$). For verbal aggression, survival rates differed significantly between those rated low and moderate risk ($\chi^2 = 7.651$, $df=1$, $p<.01$), low and high risk ($\chi^2 = 12.202$, $df=1$, $p<.001$), but

was not significantly different between those rated moderate and high risk ($\chi^2=1.089$, $df=1$, $p=.297$). For self-harm, survival rates differed significantly between groups rated as low and moderate risk for self-harm ($\chi^2 = 24.943$, $df=1$, $p<.001$), and between low and high risk groups ($\chi^2 = 33.290$, $df=1$, $p<.001$), but did not differ significantly between moderate and high risk groups ($\chi^2 = 1.591$, $df=1$, $p=.207$). Finally, for victimisation, survival function differed significantly between moderate and high rated groups ($\chi^2 = 18.292$, $df=1$, $p<.001$), low and high rated groups ($\chi^2 = 20.183$, $df=1$, $p<.001$), and between low and moderate rated groups ($\chi^2 = 18.292$, $df=1$, $p<.001$).

Finally, the hazard function $h(t)$ for each outcome-SRE pairing at each temporal gate was calculated. Declining $h(t)$ at gate 3 suggested START still had some utility in predicting newly violent and aggressive patients 3-months after assessment (see Table 2). Similarly, the START still had utility at 3-months for the prediction of new cases of victimisation. For self-harm outcomes the hazard function increased by the third temporal gate suggesting that reassessment may be warranted earlier. The predictive utility that the START Strengths scale demonstrated for predicting self-neglect was likely to be better over a shorter (2-month) period than current guidelines suggest.

DISCUSSION

The current study has demonstrated in an adequately powered sample that elements of the START significantly predict aggressive, self-harm, and victimisation outcomes over 1-, 2- and 3-month periods. The overall result replicates that in our previous analysis (O'Shea *et al.*, 2015); however, the current study aimed to address a more nuanced question about the appropriate interval before reassessment. First, analysis using ROC principles demonstrated that AUC values

for violent and self-harm outcomes were mostly significant at all three temporal gates where previously only outcome at 3-months was known. The SRE victimisation predicted its associated outcome at 3-months only; and the Strengths scores for prediction of self-neglect were significant predictors at gates 1 and 2 only. This latter finding revealed a predictive utility not detected in our previous analysis.

Survival analysis suggested differences in time to outcome failure between groups designated at different levels of risk for each outcome largely in the expected manner; for aggressive and self-harm outcomes patients rated at different levels of risk survived (i.e. did not engage in the adverse behaviour) for significantly different periods following assessment. For victimisation, those at high risk survived significantly less well, while those at low and medium risk did not differ in their survival rates. This suggests that those rated as non-high risk in this group may not warrant different preventive interventions. Finally, hazard function analysis suggested that, for aggressive and victimisation outcomes, the initial assessment may still have predictive value at the three month temporal gate; however, that for self-harm is unlikely to distinguish between groups beyond the 2-month gate.

Practice implications and considerations for future research

The current study provides the first evidence about the appropriate interval before reassessment with the START, a tool that is growing in popularity and which has been used in a number of countries (O'Shea & Dickens, 2014). Our findings have important implications for risk assessment practice because they suggest that different reassessment intervals are appropriate for the various START outcomes. However, we do not recommend that START raters begin to conduct rating of different elements of the tool at different intervals as this would

prove overly complex. A balanced decision should be arrived at about the appropriate reassessment interval based on the totality of the evidence. Hazard function analysis suggested that 3- month assessment is adequate for reassessment of physical aggression and victimisation outcomes. However, given that 74 (46.8%) of the patients rated as low or moderate risk for physical aggression engaged in the outcome by gate 2 suggests that these patients would, in any event, warrant reassessment since it may be necessary to adjust their SRE upward. Further, nine (22.5%) of the patients rated high risk, and 28 (38.9%) of the patients rated as moderate risk, had not engaged in aggression by gate 3 and thus may have warranted reassessment to a lower risk status at an earlier point. Thus, in over half the cases (51.2%), an earlier reassessment at 2 months might have improved accuracy, or prevented unnecessarily restrictive risk management interventions. Such a magnitude would not have been the case should the reassessment have taken place at the 1- month gate. Hazard function analysis suggested that optimum time for reassessment for self-harm was at the 2-month gate which further supports our case. Finally, there appears to be significant predictive validity for the Strength score for self-neglect up to 2-months. As a result, we think that raters should bear this finding strongly in mind when considering a SRE for this outcome as this should result in greater accuracy. Based on the totality of evidence we think that clinicians using the START could consider amending their practice to reassess every 60 rather than every 90 days; or routinely revisit the assessment when the predicted outcome occurs before the recommended 90 day interval. We do not believe this recommendation is too onerous given that start assessment generally takes between 25 and 40 minutes (Desmarais et al., 2011) and the mean time of completion decreases on subsequent assessments (Quinn et al., 2013).

Future research should aim to replicate findings of the current study with other risk assessment tools used in clinical practice to determine the most appropriate reassessment interval. Further work is required to determine the reasons for the self-harm SRE's outperformance of the Strength and Vulnerability scores. If the factors that raters are considering in their successful prediction of self-harm can be determined then these could be considered for inclusion among the START items; this should result in better estimates overall.

Limitations

It is possible that, given outcomes data beyond that available to us, hazard function analysis might have suggested that START might be warranted over periods longer than 3-months for aggressive outcomes. We did not collect data to test this. However, we think that our overall conclusion would be unlikely to change given the survival attrition at temporal gates 1 and 2. The study employed a pseudo prospective design where risk management is the responsibility of those who also record outcomes in the clinical record. However, clinicians were not aware the data was being collected for analysis, and raters were independent of assessors at the point of coding. Assessment was undertaken as part of routine clinical practice rather than by independent researchers, and it is possible there is variation between rating teams; however, this is the recommended procedure for completing START assessment. Length of stay of patients was highly variable; this was due to our using the first iteration of the START following its introduction across the study setting including with existing patients. Further research with a larger sample would be required to ascertain whether survival is moderated by length of stay. Finally, our calculations are based on the occurrence of adverse behaviours meeting or exceeding level 1 criteria described in the SOS; it is therefore possible that an analysis of incidents with

more serious outcomes would produce different results. However, it is well documented that incident outcome in inpatient care is not always directly related to the intention and commitment of the person displaying the behaviour (e.g., Bowers et al., 2007). Finally, it is possible that there could have been under-reporting of incidents although we have found rates in the current study comparable to, or exceeding, those reported elsewhere (O'Shea *et al.*, In Press).

Summary and conclusions

The current study has examined the most appropriate interval for START reassessment relating to a number of adverse outcomes. The current 90-day recommendation is supported by the evidence of this study. However, we think that the START assessment should be routinely revisited if target behaviours occur in the shorter term, or where they do not occur in individuals rated as at elevated risk by 60 days. This should provide the opportunity to more accurately estimate risk and balance management interventions accordingly in order to protect staff and patients from unwanted behaviours, and minimize restrictive practices on an individual level. In secure and forensic mental health settings, nurses are the professionals best equipped to lead on day-to-day risk assessment and management; the use of the START at appropriate intervals can assist them.

REFERENCES

- Bowers, L., Nijman, H. , Palmstierna, T. and Crowhurst, N. (2007). Issues in the measurement of violent incidents and the introduction of a new scale: the 'attacks' (attempted and actual assault scale). *Acta Psychiatrica Scandinavica*, **106**: s142, 106-109.
- Chu, C. M., Thomas, S. D. M., Ogloff, J. R. P., & Daffern, M. (2013). The short- to medium-term predictive accuracy of static and dynamic risk assessment measures in a secure forensic hospital. *Assessment*, **20**, 230–241. doi:10.1177/1073191111418298
- Crocker, A. G., Braithwaite, E., Laferriere, D., Gagnon, D., Venegas, C., & Jenkins, T. (2011). START changing practice: Implementing a risk assessment and management tool in a civil psychiatric setting. *The International Journal of Forensic Mental Health*, **10**, 13–28. doi: 10.1080/14999013.2011.553146
- Crocker, A. G., Garcia, A., Israel, M., Hindle, Y., Gagnon, D., & Venegas, C. (2008). *Implementing and using a systematic risk assessment scheme to increase patient safety on a risk management unit for individuals with severe mental illness: A demonstration project*. Edmonton, Canada: Canadian Patient Safety Institute.
- Desmarais, S. L., Collins, M., Nicholls, T., & Brink, J. (2011). Perceptions of the Short-Term Assessment of Risk and Treatability (START) as implemented in forensic psychiatric practice. Unpublished manuscript.
- Dickens, G.L. & Doyle, M. (In Press). Mentally disordered firesetters in secure mental health care: a forensic mental health nursing perspective. In: Doley, R.M., Dickens, G.L., & Gannon, T.A. (Eds) *The psychology of arson: a practical guide to understanding and managing adult deliberate firesetters*. London: Routledge.
- Dolan, M., & Doyle, M. (2000). Violence risk prediction. *The British Journal of Psychiatry*, **177**, 303-311.
- Doyle, M. & Jones, P. (2013). Hodges' Health Career Model and its role and potential application in forensic mental health nursing. *Journal of Psychiatric and Mental Health Nursing*, **20**, 631-640. doi: 10.1111/j.1365-2850.2012.01961.x
- Doyle, M., Lewis, G., & Brisbane, M. (2008). Implementing the Short-Term Assessment of Risk and Treatability (START) in a forensic mental health service. *Psychiatric Bulletin*, **32**, 406–408. doi:10.1192/pb.bp.108.019794
- Fleischman, R.J., Adams, A.L., Hedges, J.R., John Ma, O., Mullins, R.J., & Newgard, C.D. (2010). The optimum follow-up period for assessing mortality outcomes in injured elders. *Journal of the American Geriatric Society*, **58**, 1843-1849. doi.org/10.1111%2Fj.1532-5415.2010.03065.x
- Hart, S.D., Kropp, P.R., & Laws, D.R., Klaver, J., Logan, C., & Watt, K.A. (2003). *The Risk for Sexual Violence Protocol (RSVP)*. Burnaby, BC: Simon Fraser University, Mental Health, Law, and Policy Institute.

- Nicholls, T. L., Gagnon, N., Crocker, A. G., Brink, J., Desmarais, S. L., & Webster, C. (2007). *START Outcomes Scale (SOS)*. Vancouver: BC Mental Health & Addiction Services.
- O'Shea, L.E. & Dickens, G.L. (2014). Short-Term Assessment of Risk and Treatability (START): Systematic review and meta-analysis. *Psychological Assessment*, **26**, 990-1002.
- O'Shea, L.E., Picchioni, M.M., & Dickens, G.L. (2015) The predictive validity of the Short-Term Assessment of Risk and Treatability (START) for multiple adverse outcomes. *Assessment*. doi:10.1177/1073191115573301
- O'Shea, L.E. & Dickens, G.L. (2015). Predictive validity of the START for unauthorized leave and substance abuse in a secure mental health setting. *International Journal of Nursing Studies*. doi:10.1016/j.ijnurstu.2015.02.007
- Ogloff, J.R.P., & Daffern, M. (2006). The Dynamic Appraisal of Situational Aggression: An instrument to assess risk for imminent aggression in psychiatric inpatients. *Behavioral Sciences and the Law*, **24**, 799-813.
- Quinn, R., Miles, H., & Kinane, C. (2013). The Validity of the Short-Term Assessment of Risk and Treatability (START) in a UK medium secure forensic mental health service. *The International Journal of Forensic Mental Health*, **12**, 215–224. doi:10.1080/14999013.2013.832714
- van den Brink, R., Troquete, N., van Os, T., Schaafsma, G., Schram, A., & Wiersma, D. (2009). *Patient self-appraisal of the risk and protective factors of the START*. Paper presented at the 9th Annual Conference of the International Association of Forensic Mental Health Services, Edinburgh, United Kingdom.
- Webster, C., Douglas, K., Eaves, D., & Hart, S. (1997). *HCR-20: Assessing risk of violence (Version 2)*. Burnaby, British Columbia, Canada: Simon Fraser University, Mental Health, Law, and Policy Institute.
- 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)*. Coquitlam, Canada: British Columbia Mental Health & Addiction Services.
- Wilson, C. M., Desmarais, S. L., Nicholls, T. L., & Brink, J. (2010). The role of client strengths in assessments of violence risk using the Short-Term Assessment of Risk and Treatability (START). *The International Journal of Forensic Mental Health*, **9**, 282–293. doi:10.1080/14999013.2010.534694
- Wilson, C. M., Desmarais, S. L., Nicholls, T. L., Hart, S. D., & Brink, J. (2013). Predictive validity of dynamic factors: Assessing violence risk in forensic psychiatric inpatients. *Law and Human Behavior*, **37**, 377–388. doi:10.1037/lhb0000025
- World Health Organisation. (1992). *The ICD-10 Classification of Mental and Behavioural disorders*. Geneva: WHO.

Table 1: START (Webster et al., 2009) items

Social skills
Relationships
Occupational
Recreational
Self-care
Mental state
Emotional state
Substance use
Impulse control
External triggers
Social support
Material resources
Attitudes
Medication adherence
Rule adherence
Conduct
Insight
Plans
Coping
Treatability

Table 2: Hazard function [$h(t)$] analysis for SREs at three temporal gates

SRE	Temporal Gate								
	Month 1			Month 2			Month 3		
	<i>n</i> failing cum. n (%)	<i>N</i> potential survivors	<i>h</i> (<i>t</i>)	<i>n</i> failing cum. n (%)	<i>N</i> potential survivors	<i>h</i> (<i>t</i>)	<i>n</i> failing cum. n (%)	<i>N</i> potential survivors	<i>h</i> (<i>t</i>)
<i>Physical aggression</i>									
All (<i>N</i> =198)	74 (37.4)	198	.374	88 (44.4)	123	.114	105 (53.0)	110	.155†
Low risk (<i>n</i> =86)	16 (18.6)	86	.186	23 (26.7)	63	.111	30 (34.9)	63	.111†
Moderate risk (<i>n</i> =72)	31 (43.1)	72	.431	38 (52.8)	38	.200	44 (61.1)	34	.176†
High risk (<i>n</i> =40)	27 (67.5)	40	.675	27 (67.5)	29	.184	31 (77.5)	13	.153†
<i>Verbal aggression*</i>									
All (<i>N</i> =198)	84 (46.2)	198	.462	111 (56.1)	87	.310	116 (59.8)	82	.061†
Low risk (<i>n</i> =86)	24 (27.9)	86	.279	37 (43.0)	49	.265	40 (47.6)	46	.065†
Moderate risk (<i>n</i> =72)	38 (52.8)	72	.528	47 (65.3)	25	.259	47 (65.3)	25	-
High risk (<i>n</i> =40)	24 (60.0)	40	.611	27 (67.5)	13	.231	29 (74.4)	11	.182†
<i>Selfharm</i>									
All (<i>N</i> =163)	32 (19.6)	163	.196	39 (23.9)	131	.053	45 (27.6)	124	.048†
Low risk (<i>n</i> =103)	8 (7.8)	103	.078	10 (9.7)	95	.021†	12 (11.6)	85	.023
Moderate risk (<i>n</i> =38)	11 (28.9)	38	.289	15 (39.5)	27	.148†	19 (50.0)	12	.333
High risk (<i>n</i> =22)	13 (59.1)	22	.591	14 (63.6)	9	.111†	14 (63.6)	7	.714
<i>Victimisation</i>									
All (<i>N</i> =189)	16 (8.5)	189	.085	27 (14.3)	173	.064	35 (18.5)	162	.049†
Low risk (<i>n</i> =113)	8 (7.1)	113	.071	14 (12.4)	105	.057	15 (13.3)	99	.010†
Moderate risk (<i>n</i> =49)	1 (2.0)	49	.020	2 (4.1)	48	.125	6 (12.2)	47	.085†
High risk (<i>n</i> =27)	7 (25.9)	27	.259	11 (40.7)	20	.200	14 (51.9)	16	.188†
<i>Selfneglect</i>									
All (<i>N</i> =185)	17 (9.2)	185	.092	24 (13.0)	168	.042†	33 (17.8)	161	.056
Low risk (<i>n</i> =108)	10 (9.3)	108	.093	13 (12.0)	98	.031†	17 (15.7)	105	.038
Moderate risk (<i>n</i> =54)	2 (3.7)	54	.037	5(9.3)	52	.058†	10 (18.5)	49	.102
High risk (<i>n</i> =23)	5 (21.7)	23	.217	6 (26.1)	18	.056	6 (26.1)	17	-

$h(t)$ =hazard function calculated as number failing to survive as a proportion of all surviving cases. SRE = Specific Risk Estimate for associated outcome except * (SRE for physical aggression)

† indicates last temporal gate in the analysis at which $h(t)$ declines indicating optimum time for reassessment.

Table 3: AUC values for Strength and Vulnerabilities scores and SREs at three temporal gates

	Temporal Gate		
	1 Month	2 Months	3 Months
	AUC (95% CI)	AUC (95% CI)	AUC (95% CI)
<i>Physical Aggression</i>			
Strengths	.626 (.547, .704)**	.665 (.591, .740)***	.653 (.576, .731)***
Vulnerabilities	.600 (.521, .679)*	.583 (.504, .662)*	.614 (.535, .693)**
SRE Aggression	.713 (.638, .787)***	.678 (.602, .753) ***	.682 (.608, .757) ***
<i>Verbal Aggression</i>			
Strengths	.644 (.566, .721)**	.655 (.578, .732)***	.688 (.611, .766)***
Vulnerabilities	.600 (.522, .679)*	.569 (.487, .650)	.574 (.490, .658)
SRE Aggression	.648 (.571, .725)*	.618 (.539, .697)*	.622 (.542, .702)**
<i>Selfharm</i>			
Strengths	.579 (.467, .690)	.582 (.473, .691)	.578 (.473, .682)
Vulnerabilities	.539 (.428, .649)	.537 (.428, .646)	.551 (.446, .656)
SRE self-harm	.768 (.669, .867)***	.775 (.678, .871)***	.788 (.700, .877)***
SRE suicide	.654 (.530, .779)*	.668 (.554, .783)**	.691 (.583, .799)**
<i>Victimisation</i>			
Strengths	.609 (.470, .747)	.541 (.423, .658)	.605 (.509, .702)
Vulnerabilities	.590 (.453, .727)	.592 (.482, .701)	.567 (.466, .668)
SRE Victimisation	.611 (.442, .779)	.602 (.470, .734)	.653 (.540, .765)**
<i>Self neglect</i>			
Strengths	.691 (.572, .811)*	.691 (.593, .789)*	.630 (.528, .732)
Vulnerabilities	.610 (.452, .769)	.619 (.488, .750)	.566 (.446, .686)
SRE Self neglect	.615 (.446, .785)	.605 (.453, .756)	.546 (.410, .681)

* P<.05 ** P<.01 *** P<.001







