Published in final edited form as:

Psychol Assess. 2012 September ; 24(3): 685–700. doi:10.1037/a0026668.

Using Dynamic Risk and Protective Factors to Predict Inpatient Aggression: Reliability and Validity of START Assessments

Sarah L. Desmarais¹, Tonia L. Nicholls^{2,3,4}, Catherine M. Wilson^{2,3}, and Johann Brink^{2,3,4}

¹Department of Mental Health Law & Policy and Department of Community & Family Health, University of South Florida

²Forensic Psychiatric Hospital, BC Mental Health & Addiction Services Commission, Department of Psychiatry, University of British Columbia, and Department of Psychology, Simon Fraser University

³Department of Psychology, Simon Fraser University, and Forensic Psychiatric Hospital, BC Mental Health & Addiction Services Commission

⁴Forensic Psychiatric Hospital, BC Mental Health & Addiction Services Commission, Department of Psychiatry, University of British Columbia, and Department of Criminology, Simon Fraser University

Abstract

The Short-Term Assessment of Risk and Treatability (START) is a relatively new structured professional judgment guide for the assessment and management of short-term risks associated with mental, substance use, and personality disorders. The scheme may be distinguished from other violence risk instruments because of its inclusion of 20 dynamic factors that are rated in terms of both vulnerability and strength. This study examined the reliability and validity of START assessments in predicting inpatient aggression. Research assistants completed START assessments for 120 male forensic psychiatric patients through review of hospital files. They additionally completed Historical-Clinical-Risk Management – 20 (HCR-20) and the Hare Psychopathy Checklist: Screening Version (PCL:SV) assessments. Outcome data was coded from hospital files for a 12-month follow-up period using the Overt Aggression Scale (OAS). START assessments evidenced excellent interrater reliability and demonstrated both predictive and incremental validity over the HCR-20 Historical subscale scores and PCL:SV total scores. Overall, results support the reliability and validity of START assessments, and use of the structured professional judgment approach more broadly, as well as the value of using dynamic risk and protective factors to assess violence risk.

Keywords

START; violence risk assessment; structured professional judgment; inpatient aggression; dynamic factors; protective factors; HCR-20; PCL:SV

Correspondence concerning this article should be addressed to Sarah L. Desmarais, University of South Florida, 13301 Bruce B. Downs Blvd (MHC 2735), Tampa, FL 33612. sdesmarais@usf.edu.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/pas

The notion of increasing structure to improve decision making is hardly new, either generally or as it relates to assessing violence risk specifically (Scott, 1977). Despite a professional commitment to evidence-based practice in psychology (American Psychological Association, 2006), unstructured psychological assessment is an approach commonly used by clinicians to estimate risk of future violence (Higgins, Watts, Bindman, Slade, & Thornicroft, 2005). Yet, the research evidence demonstrates that unstructured assessments of violence risk are both less reliable and less valid than structured approaches (Grove, Zald, Lebow, Snitz, & Nelson, 2000). Unfortunately, such assessments may over- or underestimate risk, resulting in potentially serious consequences for public health and safety, as well as for clinical outcomes.

As a result of concerns regarding the accuracy of unstructured clinical approaches to violence risk assessment, many different technologies to assist clinicians in the process of assessing and managing violence risk have been introduced over the past few decades; prominent among these are actuarial and structured professional judgment instruments. Various structured professional judgment guides, such as the Historical-Clinical-Risk Management – 20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997), have been developed to guide assessments around a set of empirically-identified and theoretically-sound factors, but also in accord with best practice clinical standards. Whereas actuarial instruments, such as the Violence Risk Appraisal Guide (VRAG; Quinsey, Harris, Rice, & Cormier, 2006), the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995), and the Classification of Violence Risk (COVR; Monahan et al., 2005), take a mechanical and algorithmic prediction approach to predicting violence risk (Grove & Meehl, 1996), structured professional judgment guides are designed to structure clinicians' assessments of violence risk while still allowing for consideration of client- or context-specific issues (Douglas & Kropp, 2002). Though there remains considerable debate in the violence risk assessment field regarding the superiority of one approach over the other, recent meta-analyses show that both approaches improve violence risk assessment accuracy over unstructured assessments at comparable rates (Campbell, French & Gendreau, 2009; Guy, 2008).

The introduction of structured approaches represents a significant advance in the science and practice of violence risk assessment; however, important limitations remain (Hanson, 2005; Heilbrun, Yasuhara, & Shah, 2010; Rogers, 2000). First, existing instruments emphasize static and stable factors, such as patient sex or history of violence. The predictive capacity of such variables is well-established (see Campbell et al., 2009 for a meta-analytic review); however, they offer little in the way of guidance regarding clinical intervention (Douglas & Skeem, 2005). Though they may inform decision making regarding levels of supervision or community access (Hart, Webster, & Douglas, 2001), they fail to identify targets for therapeutic intervention. In contrast, dynamic factors may be amenable to treatment and, when targeted during treatment, may contribute to reductions in violence risk (Douglas & Skeem, 2005). Recent research provides evidence for both the independent and incremental validity of dynamic factors over static factors in the assessment of future violence risk (Brown, Amand, & Zamble, 2009; Chu, Thomas, Ogloff, & Daffern, 2011; Gagliardi, Lovell, Peterson, & Jemelka, 2004; McDermott, Edens, Quanbeck, Busse, & Scott, 2008; Simourd, 2004; Wilson, Desmarais, Nicholls, Hart, & Brink, under review), although their utility remains controversial (Philippe, Koeter, van der Staak, & van den Brink, 2006; Rice, Harris, & Quinsey, 2002).

Second, the traditional focus of violence risk assessment research has been on risk factors; that is, those factors that increase the likelihood that an individual will engage in future violence. There is very little research on factors that would protect against or decrease violence risk among adults, despite criticism that focus on risk factors may produce biased assessments that likely overestimate risk (e.g., Rogers, 2000; Ryba, 2008; Webster,

Nicholls, Martin, Desmarais, & Brink, 2006). The adolescent offender literature provides support for the inclusion of protective factors in the violence risk assessment process. In their evaluation of the role of resources in models of risk, for example, Gilgun and colleagues (2000) found that inclusion of resources significantly increased the fit of predictive models and increased the classification accuracy for adolescent inmate ($n = 1,311$) and comparison ($n = 1,702$) samples over risk-only models. With limited empirical evidence in the adult violence risk assessment literature, the role of protective factors remains subject to debate. However, their inclusion in the risk assessment process may have important therapeutic benefits: Attending to positive attributes may foster the therapeutic relationship, increase patient engagement, and contribute to establishing treatment goals (Duckworth, Steen, & Seligman, 2005; Saleebey, 1996). By extension, consideration of protective factors may improve violence risk management efforts via these and other mechanisms. The Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Middleton, 2004; Webster, Martin, Brink, Nicholls, & Desmarais, 2009) provides a framework for structuring assessments of dynamic protective and risk factors.

Short-Term Assessment of Risk and Treatability

START guides a multi-faceted, idiographic assessment of short-term risks associated with mental, substance use, and personality disorders. In contrast with existing instruments that comprise *either* risk (e.g., HCR-20; Webster et al., 1997) *or* protective factors (e.g., the Structured Assessment of Protective Factors, SAPROF, de Vogel, de Ruiter, & de Vries Robbé, 2009), the START incorporates *both* client strengths *and* vulnerabilities in one comprehensive assessment. Designed primarily as a clinical assessment and treatment-planning guide (recognizing that assessments may be aggregated for research purposes; Webster et al., 2009), START also may inform administrative tasks, such as identification of treatment programs, monitoring of patient progress, and service-wide surveillance (Desmarais, Webster, Martin, Dassinger, Brink, & Nicholls, 2007). START was recently identified as a promising new instrument for assessing risk for aggression in psychiatric inpatient settings (Daffern, 2007), as well as a tool for supporting best practice in managing violence and related risks (Department of Health, 2007; Haque, Cree, Webster, & Hasnie, 2008). A copy of the START Summary Sheet is provided in the Appendix.

START has experienced quick uptake into clinical practice (Webster et al., 2009): We are aware of implementations in at least 10 countries, and the manual has been translated into four languages with an additional four translations underway. However, only a handful of studies have examined the reliability and validity of START assessments (Braithwaite, Charette, Crocker, & Reyes, 2010; Chu et al., 2011; Gray et al., 2011; Nicholls, Brink, Desmarais, Webster, & Martin, 2006; Nonstad et al., 2010). The original validation study was published by Nicholls and colleagues (2006). Evaluating START assessments completed by nurses, social workers, and psychiatrists regarding 137 male forensic psychiatric patients, Nicholls et al. found excellent interrater agreement overall (intraclass correlation coefficient, $ICC_2 = .87, p < .001$) and within professional disciplines (nursing: .88; social work: .92; and psychiatry: .80, all $p < .001$). The authors also reported significantly higher START total scores¹ for patients who engaged in aggression over the 12-month follow-up: any aggression to others ($M = 75.66$ vs. 65.86), verbal aggression ($M = 75.86$ vs. 66.82), aggression against objects ($M = 77.90$ vs. 68.00), physical aggression against others ($M = 76.32$ vs. 68.25), violence against others ($M = 81.82$ vs. 69.12), and

¹At the time of the original validation study, START items were rated on one continuous 6-point scale, from 0 indicating considerable strength to 5 indicating considerable vulnerability, and included final risk estimates for only four outcome domains (violence to others, self-harm, suicide, and unauthorized absence).

sexual aggression ($M = 80.63$ vs. 70.24), all $p < .05$. Receiver Operating Characteristic (ROC) analyses of a subsample of 50 patients who remained hospitalized throughout follow-up revealed good validity in predicting verbal aggression ($AUC = .72$, $SE = .07$, $p < .001$), physical aggression against objects ($AUC = .67$, $SE = .08$, $p < .05$), physical aggression against others ($AUC = .70$, $SE = .08$, $p < .01$), and sexually inappropriate behavior ($AUC = .92$, $SE = .10$, $p < .05$).

Results of this research are promising, but further evaluation is necessary for several reasons. The predictive validity of the final risk estimates, one of the identifying features of the structured professional judgment approach (Skeem & Monahan, 2011; Singh, Grann, & Fazel, 2011), has only been examined in two studies (Braithwaite et al., 2010; Gray et al., 2011) and the validation samples have been quite small (n 's = 34 – 50). Furthermore, there have been several significant changes to the instrument since these evaluations. Now, each of the 20 START items is scored in terms of both vulnerability and strength, and final risk estimates of low, moderate, or high are made across seven outcome domains (Webster et al., 2004). In 2009, the authors published a text revision of the START manual (Version 1.1; Webster et al., 2009), which included content updates to three items (Mental State, Emotional State, and Treatability), explicit operationalization of START components left undefined in the 2004 consultation edition of the manual (e.g., each of the risk outcome domains, strengths and vulnerabilities, and key and critical items), and specification of coding time frames (i.e., item ratings based on functioning over the past two to three months or since the previous START assessment).

A recent analysis of 30 male forensic psychiatric inpatients provided evidence for the ability of assessments completed using START Version 1.1 to predict short-term violence risk (Wilson et al., 2010). Specifically, results demonstrated that Strength ($AUC = .73$, $SE = .05$, $p < .001$) and Vulnerability ($AUC = .74$, $SE = .05$, $p < .001$) total scores as well as the final risk estimates predicted aggressive behavior in the three months following the assessment ($AUC = .82$, $SE = .05$, $p < .001$). Findings also supported the dynamic nature of the START items: survival analyses revealed that dynamic factors evidenced change over time, and that it was the change in scores that was most predictive of future aggression (Wilson et al., under review). However, given the small sample size, more research is needed to establish the reliability and validity of assessments completed using START in its current form.

The Present Study

The present study extends the work of Wilson and colleagues, increasing the sample size to 120 patients and focusing on a 12-month follow-up period, allowing for a more robust and nuanced analysis of the reliability and validity of START assessments in predicting inpatient aggression. We had three specific research questions: (1) What is the interrater reliability of START assessments? (2) Do START assessments predict various forms of inpatient aggression? (3) Do START assessments show incremental validity in predicting aggression over assessments of historical risk factors?

Methods

Participants

Assessments were completed for 120 male patients in a secure forensic psychiatric hospital in western Canada. These participants were randomly selected from the population of 527 patients who participated in a study conducted by Verdun-Jones and colleagues (2006) which examined the prevalence and severity of inpatient aggression at this hospital over a one-year period (from January 1 to December 31, 2004). Most participants selected for inclusion in the present study were diagnosed with schizophrenia spectrum disorders

(85.0%, $n = 102$) and often had co-morbid substance use disorders (52.5%, $n = 63$). The vast majority had prior charge(s) (75.0%, $n = 90$) and previous mental health contact(s) (92.5%, $n = 111$) before the index offence. Index offences were predominantly violent in nature (80.0%, $n = 96$). Average length of patient stay prior to the beginning of the study period was approximately six years ($M = 2235.93$ days, $SD = 2260.08$) and mean subject age was 37.97 years ($SD = 11.74$). Breakdown in terms of subject race/ethnicity was as follows: 75.8% ($n = 91$) White, 10.0% ($n = 12$) First Nations, 3.3% ($n = 4$) Asian, 2.5% ($n = 3$) Black, and 6.7% ($n = 8$) 'other'. Eighty-nine percent of participants ($n = 107$) were in hospital as a result of having been found not criminally responsible on account of mental disorder (i.e., Canada's insanity defense; Desmarais, Hucker, Brink & De Freitas, 2008), 5.0% ($n = 6$) were on remand awaiting assessment, 3.3% ($n = 4$) had been found unfit to stand trial (i.e., akin to the US determination of incompetent to stand trial; O'Shaughnessy, 2007), and 2.5% ($n = 3$) had been committed involuntarily.

Measures

Short-Term Assessment of Risk and Treatability (START; Webster et al., 2009)

—The START is a structured professional judgment guide for the assessment of seven often inter-related risks associated with mental, substance use, and personality disorders in adults: violence to others, self-harm, suicide, unauthorized leave, substance abuse, self-neglect and being victimized. The instrument consists of 20 dynamic factors that are assessed for both Strength and Vulnerability on a 3-point ordinal scale from 0 (minimally present) to 2 (maximally present). Strength and Vulnerability ratings should be scored independent of one another, and a patient may be scored high (or low) on both Strength and Vulnerability for any particular item. For example, a patient may receive a high Vulnerability rating for Relationships (item 2) if they are involved in an abusive intimate relationship, but also may receive a high Strength rating if s/he has a warm, loving, and reciprocal relationship with his or her parents, other family members, or peers. Based on item ratings, identification of key and critical items (i.e., items that are particularly relevant, either recently or historically, to individual risk), and consideration of historical factors, assessors estimate risk as low, moderate, or high for each of the seven outcome domains. Strength and Vulnerability total scores can be calculated for research purposes by summing the item ratings (possible range = 0 – 40). START is intended for use with both inpatient and outpatient populations in civil psychiatric, forensic psychiatric, and correctional settings.

Historical-Clinical-Risk Management – 20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997)

—The HCR-20 is a structured professional judgment guide that consists of 20 risk factors distributed across three subscales: Historical (10 items), Clinical (5 items), and Risk Management (5 items). Each item is rated on a 3-point ordinal scale from 0 (not present) to 2 (definitely present). Based on the item ratings and after considering the relevance of each factor to the individual being assessed as well as the context (e.g., hospital vs. community), the assessor arrives at a final estimate of risk for future violence (low, moderate, or high). The item ratings can be summed for research purposes for an overall total score and for each subscale. The HCR-20 is one of the most (if not the most) widely used structured professional judgment guides for assessing violence risk, translated into 16 languages and adopted by numerous mental health, forensic, and criminal justice agencies in the United States, Canada, the United Kingdom, Europe, Asia, Australia, and New Zealand.

The reliability and validity of the HCR-20 have been evaluated in over a hundred studies (Douglas, Guy, Reeves, & Weir, 2008). A recent meta-analysis found mean weighted effect sizes (AUCw) ranging from .62 ($SE = .05$) to .79 ($SE = .05$) for the violence risk estimate, total score, and the Historical, Clinical, and Risk Management subscales in predicting violence or physical aggression (Guy, 2008). Extant research suggests acceptable reliability

(values typically $>.70$) across civil psychiatric, forensic psychiatric, and correctional settings (see Douglas, Guy, Reeves, & Weir, 2008). In the current study, interrater reliability for the HCR-20 assessments ranged from good (total score: $ICC_2 = .71$) to excellent (C score: $ICC_2 = .88$), all $p < .001$.

Hare Psychopathy Checklist: Screening Version (PCL:SV; Hart, Cox, & Hare, 1995)—The PCL:SV is a 12-item measure for the assessment of psychopathic traits. Each item is rated 0 (not present), 1 (possibly present), or 2 (definitely present), and ratings are summed to arrive at a total score. A cutoff score of 18 typically is used for classification of psychopathy. The PCL:SV's conceptualization of psychopathy is analogous to that of the Revised Psychopathy Checklist (PCL-R; Hare, 2003) and research demonstrates excellent correspondence between the two measures in both forensic and correctional samples (Guy & Douglas, 2006). Though not designed to be a violence risk assessment instrument, the PCL:SV is frequently used to assess likelihood of violence, and research generally supports its reliability and validity in predicting community and institutional aggression (e.g., Belfrage, Fransson, & Strand, 2000; Douglas, Strand, Belfrage, Fransson, & Levander, 2005; Doyle, Dolan, & McGovern, 2002; Monahan et al., 2001; Nicholls, Ogloff, & Douglas, 2004). A recent meta-analysis found the PCL:SV produced one of the largest mean weighted effect sizes ($Z^+ = .22$) for predicting institutional violence, after the HCR-20 and LSI-R (Campbell et al., 2009). In the current study, interrater reliability for the PCL:SV was high: $ICC_2 = .77$, $p < .001$.

Overt Aggression Scale (OAS; Yudofsky, Silver, Jackson, Endicott, & Williams, 1986)—The OAS is a widely used observational measure for assessing inpatient aggression. The OAS measures the occurrence and severity of four categories of aggressive behavior: verbal aggression, physical aggression against objects, physical aggression against self, and physical aggression toward others. Each category of behavior is rated according to severity from least severe (1) to most severe (4). Interrater coding completed on a subset of 40 patients included in the aforementioned Verdun-Jones et al. (2006) study revealed reliability ranging from adequate for coding of physical aggression against objects ($ICC_2 = .67$) to excellent for coding of physical aggression toward others ($ICC_2 = .84$), all $p < .01$. In this study, we focus on the perpetration of verbal aggression, physical aggression toward objects, and physical aggression toward others.

Procedures

Two research assistants with Master's degrees in forensic psychology attended one-day workshops presented by the authors of the START, HCR-20, and PCL:SV, respectively. The workshops and subsequent training included completion of both practice cases and actual files upon which the research assistants were required to demonstrate adequate agreement with the trainers. For the purpose of the present study, assessments were completed using information available in patient hospital files up to December 31, 2003, inclusive. Each research assistant conducted assessments for approximately two-thirds (71.7%) and one-third (28.3%) of the participants, respectively. The order in which the research assistants completed the HCR-20 and START assessments was counterbalanced across participants. Interrater coding was completed on 20.0% of the total sample ($n = 24$). Research assistants were blind to the outcome data which was collected in a previous study through retrospective file review (see Verdun-Jones et al., 2006; Nicholls, Brink Greaves, Lussier, & Verdun-Jones, 2009; Lussier, Verdun-Jones, Deslauriers-Varin, Nicholls, & Brink, 2009). In this previous study, the OAS was completed regarding incidents of inpatient aggression occurring over the 12-month follow-up period (January 1 to December 31, 2004), as described above.

Data Analysis

We calculated descriptive statistics to examine the prevalence of verbal aggression, physical aggression against objects, and physical aggression toward others during the follow-up period and to investigate the psychometric properties of the START assessments. We conducted zero order correlations to examine associations between the Strength and Vulnerability ratings at the item level. Coefficients of .10 are considered small, .30 moderate, and .50 large correlations (Cohen, 1988). We evaluated interrater reliability of the Strength and Vulnerability total scores, as well as of the final violence risk estimates using ICC₂, computed using two-way mixed effects, absolute agreement models. Values greater than or equal to .75 indicate excellent agreement (Cicchetti et al., 2006). To evaluate associations between START, HCR-20 and PCL:SV scores, we calculated zero order correlations. Cohen's kappa was calculated to examine agreement between START and HCR-20 final risk estimates. Values between .00–.20 indicate slight, .21–.40 fair, .41–.60 moderate, .61–.80 substantial and .81–1.00 almost perfect agreement (Landis & Koch, 1977). We conducted *t*-test comparisons and logistic regression analyses to examine START assessments as a function of those participants who did versus did not engage in aggression during the follow-up period. We examined predictive validity of START assessments by computing AUCs. AUC values between .70–.90 indicate good predictive accuracy, and values greater than .90 indicate excellent accuracy (Swets, 1988). Finally, we conducted two sets of hierarchical logistic regression analyses to examine the incremental validity of START assessments over the HCR-20 Historical subscale scores and PCL:SV total scores, respectively. Significant chi-square change values reflect improvements in the prediction model and significant odds ratios indicate contributions of individual predictors to the overall model. All analyses were conducted using IBM SPSS Statistics 19 for Windows.

Results

The nature and severity of aggression perpetrated by participants during the study period is presented in Table 1. Almost half of the sample (45.8%, $n = 55$) remained incident free during follow-up. Consistent with prior research conducted at this institution (Nicholls et al., 2006), the most common form of aggressive behavior was verbal aggression, perpetrated by more than half of the sample (52.5%, $n = 63$), followed by physical aggression towards others (22.5%, $n = 27$), and physical aggression against objects (16.7%, $n = 20$). Despite relatively high base rates of aggression for an inpatient sample, further examination of Table 1 reveals that aggressive behaviors were generally mild to moderate in terms of severity and that severe events were rare. To demonstrate, less than 10% of participants committed physical aggression (against objects or others) in the two highest severity categories (levels 3 and 4 on the OAS); only two participants engaged in physical aggression against objects categorized as level 4; and only one participant engaged in physical aggression against others categorized as level 4. Because of the restricted severity range, data reported in subsequent analyses reflect dichotomous coding of the presence (1) or absence (0) of verbal aggression, physical aggression toward objects, and physical aggression toward others.

Descriptive statistics for the START, HCR-20, and PCL:SV scores are presented in Table 2. On the START, Strength total scores ranged from 4 to 38 ($M = 18.46$, $SD = 8.08$) and Vulnerability total scores ranged from 0 to 33 ($M = 16.82$, $SD = 8.07$), out of the possible range of 0 to 40. START final risk estimates classified approximately one-third of participants as low (38.7%), moderate (33.6%), and high (27.7%) risk for violence, respectively. A chi-square test revealed that the distribution of low, moderate, and high ratings of violence risk did not differ significantly, $\chi^2(2, N = 119) = 2.49$, $p = .288$. Similar classification rates were seen for the HCR-20 final risk judgments: 39.2% low; 33.3% moderate; and 27.5% high. The vast majority of participants (93.3%) did not meet the diagnostic criteria for psychopathy (i.e., PCL:SV ≥ 18).

Individual item scores, endorsement frequencies, and correlations between Strength and Vulnerability ratings for the START assessments are shown in Table 3. For all items, assessors made use of the full range of scores (from 0 to 2) for both Strength and Vulnerability ratings, suggesting good distribution and discrimination at the item level. For each item, associations between the Strength and Vulnerability ratings were moderate to strong and in the expected direction, ranging from $r = -.41$ for Relationships (item 2) to $r = -.84$ for Substance Use (item 8), all $p < .001$.

Strength and Vulnerability total scores were strongly correlated with violence risk estimates in the expected directions, suggesting that the item ratings were informing the final judgments of low, moderate, or high risk as intended in the structured professional judgment approach. Specifically, as Strength total scores increased, estimates of violence risk decreased ($r = -.60$), and as Vulnerability total scores increased, estimates of violence risk increased ($r = .69$), all $p < .001$. As observed at the item level, the Strength and Vulnerability total scores also were highly correlated in the expected direction, $r = -.87$, $p < .001$.

Reliability

Calculated on a subset of 24 participants, interrater reliability was high for START Strength and Vulnerability total scores, as well as for the violence risk estimates, $ICC_2 = .93$, $.95$, and $.85$, respectively, all $p < .001$.

Validity

Correlations between START, HCR-20, and PCL:SV scores are presented in Table 4 and a contingency table between START and HCR-20 violence risk estimates is presented in Table 5. Associations were moderate to strong between START Vulnerability (r range = $.31$ to $.88$) and Strength (r range = $-.43$ to $.77$) total scores and HCR-20 subscale and total scores, all $p < .001$ (see Table 4). Agreement between START and HCR-20 violence risk estimates was good, $\kappa = .77$, $p < .001$. Review of Table 5 reveals that there were no instances in which a patient was identified as high risk on one instrument and low risk on the other, or vice versa. In contrast, START Strength and Vulnerability total scores and violence risk estimates were only weakly associated with PCL:SV scores, if at all (see Table 4). Examination of ratings of those START items arguably most aligned with the construct of psychopathy, namely Impulse Control (item 9), Attitudes (item 13), and Conduct (item 16), evidenced strong associations with the Vulnerability (r range = $.40$ to $.51$) and Strength (r range = $-.40$ to $.51$) ratings, all $p < .001$.

As expected, mean START Vulnerability total scores were significantly higher among those participants who engaged in any form of aggression than those who did not ($M = 20.52$, $SD = 7.03$ vs. $M = 12.45$, $SD = 7.00$), verbal aggression ($M = 20.60$, $SD = 6.85$ vs. $M = 12.65$, $SD = 7.25$), physical aggression against objects ($M = 23.60$, $SD = 5.53$ vs. $M = 15.47$, $SD = 7.83$), and physical aggression toward others ($M = 22.74$, $SD = 7.36$ vs. $M = 15.11$, $SD = 7.46$), $t(118) = 4.42$, $d = 1.03$, $p < .001$. Similarly, mean Strength total scores were significantly higher among those participants who did *not* engage in any form of aggression ($M = 22.48$, $SD = 7.61$ vs. $M = 15.09$, $SD = 6.91$), verbal aggression ($M = 22.21$, $SD = 7.62$ vs. $M = 15.10$, $SD = 7.00$), physical aggression against objects ($M = 19.71$, $SD = 7.86$ vs. $M = 12.30$, $SD = 6.39$), and physical aggression toward others ($M = 20.26$, $SD = 7.77$ vs. $M = 12.33$, $SD = 6.01$), $t(118) = 3.96$, $d = 0.97$, $p < .001$.

Table 6 presents the percentage of participants rated as low, moderate, and high risk using the START who engaged in each type of aggressive outcome. Multinomial logistic regression analyses with estimates of high risk as the referent revealed that participants who engaged in any aggression or verbal aggression were significantly more likely to be rated as

high risk (OR = 23.02 and 18.50) than low risk, $\chi^2(2, N = 119) = 36.19$, all $p < .001$. Participants who engaged in aggression against objects were significantly more likely to be rated as high risk (OR = 5.16, $p < .01$) than moderate risk, $\chi^2(2, N = 119) = 36.19$, $p < .001$. Finally, participants who engaged in physical aggression toward others were significantly more likely to be rated as either moderate risk (OR = 69.23) or high risk (OR = 8.72) than low risk, $\chi^2(2, N = 119) = 29.38$, all $p < .001$.

AUC values for the ROC curves are presented in Table 7. START Strength and Vulnerability total scores and violence risk estimates demonstrated good predictive validity for all outcome domains, with AUC values ranging from .75 (SE = .05) for Strength total scores predicting verbal aggression to .85 (SE = .04) for violence risk estimates predicting physical aggression toward others, all $p < .001$. Review of Table 7 suggests that, in general, the predictive capacity of the START violence risk estimates exceeded that of the Strength and Vulnerability total scores, as well as that of the other instruments. For instance, the AUC values for the START violence risk estimates were several points larger than for the HCR-20 violence risk estimates and the PCL:SV total scores for any aggression (AUC = .80, SE = .04 vs. AUC = .79, SE = .04, and AUC = .75, SE = .05), verbal aggression (AUC = .78, SE = .04 vs. AUC = .74, SE = .06, and AUC = .74, SE = .05), physical aggression against objects (AUC = .84, SE = .04 vs. AUC = .70, SE = .10, and AUC = .63, SE = .07), and physical aggression toward others (AUC = .85, SE = .04 vs. AUC = .77, SE = .07, and AUC = .74, SE = .06), respectively. They also were more likely to be significant. Contrasts using z-scores, however, demonstrated that these differences in AUC values did not reach statistical significance, all $p > .05$.

To assess incremental validity of START assessments over historical risk factors, we conducted two sets of direct entry hierarchical logistic regression analyses. We first examined whether the START Strength and Vulnerability total scores and final risk estimates added to the capacity of the Historical subscale scores of the HCR-20 to predict aggression. The Historical subscale scores were added in the Step 1 of each of four models predicting any aggression, verbal aggression, physical aggression against objects, and physical aggression toward others, respectively. As may be seen in Table 8, all four models were significant. In Step 2, the Strength and Vulnerability total scores were added in one block.² Across models, the predictive capacity improved significantly; however, the models differed regarding whether the Strength or Vulnerability total scores added incremental validity (see Table 8). For any aggression and verbal aggression, the Vulnerability total scores added incremental predictive utility, whereas for physical aggression toward others, Strength total scores demonstrated incremental predictive utility. Neither Vulnerability nor Strength total scores added unique contributions to the prediction of physical aggression against objects, though the overall model was significant. In Step 3, the START violence risk estimates were added to the prediction models. For all four models, the addition produced increases in predictive capacity and revealed unique contributions of the violence risk estimates (see Table 8).

In the second set of incremental validity analyses, we examined whether the Strength and Vulnerability total scores and final risk estimates added to the capacity of the PCL:SV to predict verbal aggression, physical aggression against objects, and physical aggression towards others (see Table 9). Results were nearly identical to those found in the previous set of analyses with the HCR-20. As may be seen in Table 9, the models were significant with

²Prior to conducting these analyses, we tested for multicollinearity between the Strength and Vulnerability total scores. Specifically, we examined the variance inflation factor (VIF) derived by regressing these scores on each of the dichotomous outcomes. Despite the high correlation between the Strength and Vulnerability total scores, these diagnostic tests failed to identify multicollinearity: All VIF were < 10.

the exception of the model with physical aggression against objects as the outcome. Addition of the Strength and Vulnerability total scores in Step 2 significantly improved the predictive capacity of all models (see Table 9). Again, Vulnerability total scores evidenced unique contributions to the prediction of aggression and verbal aggression, Strength total scores evidenced unique contributions to the prediction of physical aggression toward others, and neither Vulnerability nor Strength total scores added incrementally to the prediction of physical aggression against objects. For all four models, the addition of the START violence risk estimates in Step 3 produced increases in predictive capacity and revealed unique contributions of these estimates.

Discussion

The present study adds to the emerging literature supporting the reliability and validity of START assessments. Overall, the psychometric properties of assessments completed using START Version 1.1 look strong. Although we may have anticipated ceiling effects on the Vulnerability ratings and floor effects on the Strength ratings for assessments of a relatively homogenous sample of male forensic psychiatric inpatients, this was not the case. Instead, assessors made use of the full range of possible scores for both Strength and Vulnerability ratings, as well as for the violence risk estimates. This finding suggests that START assessments may be useful for distinguishing between patients more or less likely to engage in aggressive behaviors even within a somewhat homogenous, high-risk population. Further, the results suggest that START assessments may be useful for informing case formulations and risk management strategies, and informing the allocation of scarce resources (e.g., distinguishing which patients require higher security levels).

START assessments also showed excellent interrater reliability. Though proponents of the actuarial violence risk assessment instruments often cite superiority in terms of interrater reliability as one of the principal advantages over structured professional judgment approaches (e.g., Hanson & Morton-Bourgon, 2009; Hilton, Harris & Rice, 2006), we found very high rates of agreement between raters in the present study. In fact, the degree of agreement between raters for the START Strength and Vulnerability total scores and final risk estimates met or exceeded values previously reported for leading actuarial instruments, including the VRAG (Quinsey et al., 2006), the LSI-R (Andrews & Bonta, 1995), and the PCL-R (Hare, 2003).

The strength of the interrater reliability also is noteworthy given that assessments were completed on the basis of file review alone. Past research suggests that file-based assessments consistently result in lower reliability scores than those based on multiple sources of information, such as interview plus file review (e.g., Alterman, Cacciola, & Rutherford, 1993; de Vogel & de Ruiter, 2004; Grann, Långström, Tengström, & Stålenheim, 1998; Wong, 1988). Though interviews may have contributed to even higher reliability coefficients, agreement rates were excellent even in their absence. Additionally, we may have anticipated the assessment of dynamic items and protective factors based on file information to be particularly challenging; for example, staff may be less likely to record positive or prosocial behaviors than negative or disruptive ones in a forensic mental health setting. Contrary to that expectation, agreement between raters was greater for the START Strength and Vulnerability total scores than the HCR-20 Historical subscale scores and the PCL:SV total scores which comprise historical and/or static risk factors that are fairly straightforward to code (Simourd, 2004).

Factors related to our study design may have contributed to these high rates of agreement and thus, generalizability is uncertain and findings will need to be replicated in further studies. For example, we relied on two raters, both of whom were students in the same

graduate psychology program at the time of the study. The raters also both took part in the same START training workshops. For these or other reasons, they may have been particularly like-minded and well-informed about the measure. Additionally, it is possible the files of the study hospital are particularly detailed. Further research is needed to establish the generalizability of the present reliability results and, more importantly, the “field” reliability of the START (*cf.* Edens, Boccaccini, & Johnson, 2010). We are aware of a few studies currently underway to examine the reliability (and validity) of START assessments completed by clinicians in a community clinic (Viljoen, Launeanu, Hendry, Nicholls, & Brink, 2011) and inpatient setting (Nicholls, Petersen, & Brink, 2011; Nicholls, Petersen, Brink, & Webster, 2011).

START assessments also demonstrated both predictive validity and incremental validity over the HCR-20 Historical subscale scores and PCL:SV total scores. These findings not only support the use of START for guiding assessments of risk for inpatient aggression, but also offer support for the structured professional judgment approach, more generally. In particular, the present study examined the predictive validity and incremental validity of the final risk estimates of low, moderate, or high risk, one of the hallmarks of the structured professional judgment approach (Skeem & Monahan, 2011; Singh et al., 2011). Our analyses demonstrated that these final risk estimates not only predict violence at significantly better than chance levels, but that they also add incremental predictive utility over HCR-20 Historical subscale scores, PCL:SV total scores, and START total scores. These findings contribute to a body of literature demonstrating that the final risk estimates of low, moderate, or high not only increase the clinical relevance of structured professional judgment compared to actuarial approaches to violence risk assessments, but their predictive validity as well (e.g., de Vogel & de Ruiter, 2006; Douglas, Ogloff, & Hart, 2003; Douglas, Yeomans, & Boer, 2005; for a review, see Otto & Douglas, 2010).

Our findings also add to the body of literature supporting the role of dynamic factors in risk assessment. Consistent with prior research demonstrating that dynamic factors can predict violence and recidivism as well as or better than static or highly stable risk factors (e.g., Brown et al., 2009; Gagliardi et al., 2004; McDermott et al., 2008; Simourd, 2004; Wilson et al., under review), the START Strength and Vulnerability total scores evidenced both predictive and incremental validity over HCR-20 Historical subscale scores and PCL:SV total scores. Across models, the inclusion of assessments of dynamic factors significantly reduced or eliminated the predictive capacity of assessments based on historical factors. This is not to say that historical and static factors should not figure prominently in the violence risk assessment process. Instead, “Historical variables should provide the foundation for any risk assessment” (Webster et al., 2006, p. 756) upon which assessment of dynamic factors can build to inform a contextualized and idiographic approach to violence risk management.

The correlations between the Strength and Vulnerability item ratings and total scores merit some discussion. There are different ways in which the conceptual distinction between risk and protective factors can be made. For example, some argue that they represent two poles on a continuum (Rutter, 1987), suggesting that an individual may present with considerable strengths *or* vulnerabilities on any given factor, but not both simultaneously. Conversely, as the START authors contend (Webster et al., 2006; 2009), strengths may be qualitatively different from vulnerabilities, such that a patient may be scored both with high (or low) strengths *and* vulnerabilities on a particular item. The range of *r* values observed in the present study, from $-.41$ for Relationships to $-.84$ for Substance Use, suggests that the current conceptualizations of risk and protective factors may apply differentially to particular START items. Indeed, there may be different types or categories of both risk and protective factors within the START.

The considerable overlap between the Strength and Vulnerability total scores also limits our interpretation of each as individual predictors, though it does not reduce the predictive power or reliability of the models. Though our diagnostic tests failed to identify multicollinearity between these scores, there were no models in which both the Strength and Vulnerability total scores contributed significantly. Additionally, Strength and Vulnerability total scores were found to predict different forms of aggression; specifically, Strength total scores predicted physical aggression whereas Vulnerability total scores predicted any aggression and verbal aggression. This is the first study of which we are aware to identify differential associations between Strength and Vulnerability total scores and various forms of aggression or to examine the predictive utility of both total scores within the same models. Future research is needed to determine the replicability and generalizability of these results. Taken together, these findings emphasize that our understanding of the role of protective factors in violence risk assessment is still in the very early stages.

The high degrees of association between Strength and Vulnerability item ratings and total scores also beg the question, is there value added in considering both risk and protective factors in the assessment of violence risk? We believe the answer is ‘yes’ for four reasons. First, as reviewed in the introduction of this paper, strength-informed assessments may offer a more balanced view of our patients (Rogers, 2000; Webster et al., 2006). Greater attention to positive attributes also may facilitate the establishment of therapeutic relationships which often can be challenging in forensic settings (*cf.* Wilson et al., 2010). Second, the American Psychological Association (2006) asserted that psychologists should “assess patient pathology as well as clinically-relevant strengths” (p. 276), yet this is rarely done in practice. Third, protective factors *are* significant predictors of violent outcomes. The present study and past START research demonstrate the predictive validity of Strength total scores (Braithwaite et al., 2010; Gray et al., 2011; Nonstad et al., 2010; Wilson et al., 2010). Fourth, there is indirect evidence from the present study that protective factors improve the accuracy of risk estimates: Compared to HCR-20 violence risk estimates (informed by consideration of risk factors only), the AUC values of the START violence risk estimates (informed by consideration of both risk and protective factors) were higher across all forms of aggression. Additionally, we found incremental validity for Strength total scores in the prediction of physical aggression towards others, but not verbal aggression or aggression against objects, outcomes presumably of much less concern to clinicians. This distinction is also consistent with START’s operational definition of violence risk to others, suggesting appropriate adherence to the coding anchors by the raters.

In the present study, START assessments performed as well as, and often better than, assessments completed using well-established, commonly used measures for predicting institutional aggression, namely the HCR-20 and PCL:SV. These findings are consistent with recent meta-analytic studies showing that many structured violence risk instruments (both actuarial and structured professional judgment) are more or less “interchangeable” in terms of their predictive utility (Yang, Wong, & Coid, 2010, p. 759; see also Campbell et al., 2009; Guy, 2008). Consequently, the decision regarding which instrument to use may not hinge on the psychometric superiority of assessments completed using one instrument compared to another, but rather clinical utility across other domains, such as the ability to inform management strategies or to track change over time. Recent research supports the applicability of START to these diverse clinical tasks. With respect to the former, Khiroya, Weaver, and Maden (2009) recently conducted a survey of the uptake and perceived utility of structured risk assessment measures in 29 medium secure forensic services across the United Kingdom. Though START was used relatively infrequently compared to measures that have been available in the field for considerably longer, it received the highest rating of perceived clinical utility. Implementation studies have found similar results. Specifically, clinicians using START in practice consistently report that it is a useful framework for

making judgments about risk and formulating risk management strategies (Crocker et al., 2011; Desmarais, Collins, Nicholls, & Brink, under review; Doyle, Lewis, & Brisbane, 2008; Kroppan et al., 2011). With respect to the latter and as reviewed earlier, the work of Wilson and colleagues (under review) demonstrated that START items not only were sensitive to change over time, but also that change in the dynamic item scores was predictive of future aggression. In light of the present study's positive reliability and validity results, as well as past evidence for the perceived utility and relevance to diverse clinical tasks, START appears a viable choice for clinicians interested in a structured, strength-informed approach to assess and manage risk for institutional aggression and related high-risk behaviors.

Interpretation of our findings is restricted by several limitations of the study design. First, although the START guides assessors in determining risk across seven domains, the present study measured aggressive outcomes only. The data also were restricted in the range and severity of aggressive events that occurred in the study setting, which arguably represent acts on the lower end of the violence spectrum. Thus, future studies should examine the reliability and validity of START assessments in predicting the range of adverse outcomes in populations or settings where we may anticipate a greater range of outcome severity. Second, we report findings regarding the accuracy of one-time START assessments in predicting aggression over a 12-month follow-up period. Though this design is often used in the violence risk assessment literature (Guy, 2008), the authors of the START recommend repeating assessments every two or three months (Webster et al., 2009). Unfortunately, our data do not speak to the ability of START assessments to predict aggression in the short-term (i.e., over weeks to months) and across repeated administrations. A handful of studies have examined START assessments in shorter and varied timeframes (e.g., Braithwaite et al., 2010; Chu et al., 2011; Nonstad et al., 2010; Wilson et al., 2010), but more work is needed. Third, as alluded to earlier, our study was limited by its reliance on graduate student research assistants to code all assessments on the basis of information available in file. Future studies should examine the reliability and validity of assessments informed by multiple sources of information and completed by clinicians in the "field". A fourth limitation of this study lies in our sampling of male patients admitted to one forensic psychiatric hospital. The generalizability of our findings to diverse samples and settings remains to be tested. Though studies are underway (Nicholls et al., 2011; Petersen, Douglas, & Nicholls, 2011), there is limited published research on the reliability and validity of START assessments in predicting female aggression, aggression perpetrated by outpatients and/or in community settings, as well among correction populations (e.g., inmates, parolees, probationers).

In spite of these limitations, the present study is one of the first evaluations of violence risk assessments completed using the revised version of the START and one of the only studies to include the START final risk estimates in the prediction models. These findings add essential new information to the growing evidence supporting START, and structured professional judgment more broadly, as approaches that clinicians can use to assess risk for a range of aggressive outcomes among adults with mental, substance use, and personality disorders. Findings also contribute to an emerging body of literature supporting the value of considering both risk *and* protective factors to inform assessments of violence risk. An important next step will be to examine whether consideration of dynamic risk and protective factors, and use of START in particular, improves risk management efforts and, ultimately, reduces the prevalence and severity of aggressive outcomes. Finally, a comprehensive assessment of violence risk should include consideration of service-level (e.g., staff de-escalation tactics and training, shift change) and system-level (e.g., restraint policy) factors that may increase or decrease the likelihood of patient aggression (Gadon, Johnstone, & Cooke, 2006; Harin, Iennaco, & Olsen, 2009), in addition to the more 'traditional' client-level assessment of violence risk (whether guided by the START or some other instrument).

Though there have been some recent efforts in this area (e.g., Johnstone & Cooke, 2010), continued work is needed.

Acknowledgments

This work was partially supported by Award Number P30DA028807 from the National Institute on Drug Abuse, Canadian Institutes of Health Research and Michael Smith Foundation for Health Research awards to the second author, and a Social Sciences and Humanities Research Council of Canada Graduate Scholarship to the third author.

References

- American Psychological Association Presidential Task Force on Evidence-Based Practice. Evidence-based practice in psychology. *American Psychologist*. 2006; 61:271–285. [PubMed: 16719673]
- Alterman AL, Cacciola JS, Rutherford MJ. Reliability of the Revised Psychopathy Checklist in substance abuse patients. *Psychological Assessment*. 1993; 5:442–448.
- Braithwaite E, Charette Y, Crocker AG, Reyes A. The predictive validity of clinical ratings of the Short-Term Assessment of Risk and Treatability (START). *International Journal of Forensic Mental Health Services*. 2010; 9:271–281.
- Brown SL, St. Amand MD, Zamble E. The dynamic prediction of criminal recidivism: A three-wave prospective study. *Law and Human Behavior*. 2009; 33:25–45. [PubMed: 18629620]
- Campbell MB, French S, Gendreau P. The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*. 2009; 36:567–590.
- Chu CM, Thomas SDM, Ogloff JRP, Daffern M. The short- to medium-term predictive accuracy of static and dynamic risk assessment measures in a secure forensic hospital. *Assessment*. 2011. Published online 19 August 2011.
- Cicchetti D, Bronen R, Spencer S, Haut S, Berg A, Oliver P, Tyrer P. Rating scales, scales of measurement, issues of reliability: Resolving some critical issues for clinicians and researchers. *Journal of Nervous and Mental Disorder*. 2006; 194:557–564.
- Cohen, J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum; 1988.
- Crocker AG, Braithwaite E, Laferrière, Gagnon D, Venegas C, Jenkins T. START changing practice: Implementing a risk assessment and management tool in a civil psychiatric setting. *International Journal of Forensic Mental Health*. 2011; 10:13–28.
- Daffern M. The predictive validity and practical utility of structured schemes used to assess risk for aggression in psychiatric inpatient settings. *Aggression and Violent Behavior*. 2007; 12:116–130.
- Department of Health, National Risk Management Programme. *Best practice in managing risk: Principles and evidence for best practice in the assessment and management of risk to self and others in mental health services*. London: Author; 2007.
- Desmarais, SL. START research summary. In: Webster, CD.; Martin, M-L.; Brink, J.; Nicholls, TL.; Desmarais, SL., editors. *Manual for the Short-Term Assessment of Risk and Treatability (START) (Version 1.1)*. Coquitlam, Canada: British Columbia Mental Health & Addiction Services; 2009. p. 89-104.
- Desmarais SL, Collins MJ, Nicholls TL, Brink J. Measuring clinician attitudes to bridge the gap between violence risk assessment science and practice: Perceived utility and acceptability of START. (under review).
- Desmarais SL, Hucker S, Brink J, De Freitas K. A Canadian example of insanity defence reform: Accused found not criminally responsible before and after the *Winko* decision. *International Journal of Forensic Mental Health*. 2008; 7:1–14.
- Desmarais, SL.; Webster, CD.; Martin, ML.; Dassinger, C.; Brink, J.; Nicholls, TL. *Short-Term Assessment of Risk and Treatability (START): Instructors' guide and workbook (Version 2)*. Coquitlam, Canada: Forensic Psychiatric Services Commission; 2007.
- Douglas, KS.; Guy, LS.; Reeves, KA.; Weir, J. HCR-20 violence risk assessment scheme: Overview and annotated bibliography. 2008 Nov. Retrieved June 17, 2010 from <http://kdouglas.files.wordpress.com/2006/04/annotate10-24nov2008.pdf>

- Douglas KS, Ogloff JRP, Hart SD. Evaluation of a model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services*. 2003; 54:1372–1379. [PubMed: 14557523]
- Douglas KS, Strand S, Belfrage H, Fransson G, Levander S. Reliability and validity evaluation of the Psychopathy Checklist: Screening Version (PCL:SV) in Swedish correctional and forensic psychiatric samples. *Assessment*. 2005; 12:145–161. [PubMed: 15914717]
- Douglas KS, Skeem J. Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law*. 2005; 11:347–383.
- Douglas KS, Yeomans M, Boer DP. Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior*. 2005; 32:479–510.
- Doyle M, Dolan M, McGovern J. The validity of North American risk assessment tools in predicting in-patient violent behaviour in England. *Legal and Criminological Psychology*. 2002; 7:141–154. 2002.
- Doyle M, Lewis G, Brisbane M. Implementing the Short-Term Assessment of Risk and Treatability (START) in a forensic mental health service. *Psychiatric Bulletin*. 2008; 32:406–408.
- Duckworth AL, Steen TA, Seligman MEP. Positive psychology in clinical practice. *Annual Review of Clinical Psychology*. 2005; 1:629–651.
- Edens JF, Boccaccini MT, Johnson DW. Inter-rater reliability of the PCL-R total and factor scores among psychopathic sex offenders: Are personality features more prone to disagreement than behavioral features? *Behavioral Sciences and the Law*. 2010; 28:106–119. [PubMed: 20101592]
- Gadon L, Johnstone L, Cooke D. Situational variables and institutional violence: A systematic review of the literature. *Clinical Psychology Review*. 2006; 26:515–534. [PubMed: 16828211]
- Gagliardi GJ, Lovell D, Peterson PD, Jemelka R. Forecasting recidivism in mentally ill offenders released from prison. *Law and Human Behavior*. 2004; 28:133–155. [PubMed: 15141775]
- Gilgun J, Klein C, Pranis K. The significance of resources in models of risk. *Journal of Interpersonal Violence*. 2000; 15:631–650.
- Grann M, Långström N, Tengström A, Stålenheim EG. The reliability of file-based retrospective ratings of psychopathy with the PCL-R. *Journal of Personality Assessment*. 1998; 70:416–426. [PubMed: 9760735]
- Gray NS, Benson R, Craig R, Davies H, Fitzgerald S, Huckle P, . . . Snowden RJ. The Short-Term Assessment of Risk and Treatability (START): A prospective study of inpatient behavior. *The International Journal Of Forensic Mental Health*. 2011; 10(4):305–313.
- Grove WM, Meehl PE. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*. 1996; 2:293–323.
- Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*. 2000; 12:19–30. [PubMed: 10752360]
- Guy, LS. Unpublished PhD Thesis. Burnaby, Canada: Simon Fraser University; 2008. Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey.
- Guy LS, Douglas KS. Examining the utility of the PCL:SV as a screening measure using competing factor models of psychopathy. *Psychological Assessment*. 2006; 18(2):225–230. [PubMed: 16768600]
- Hanson RK. Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence*. 2005; 20:212–217. [PubMed: 15601794]
- Hanson RK, Morton-Bourgon KE. The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis. *Psychological Assessment*. 2009; 21:1–21. [PubMed: 19290762]
- Haque Q, Cree A, Webster C, Hasnie B. Best practice in managing violence and related risks. *Psychiatric Bulletin*. 2008; 32:403–405.
- Hare, RD. *The Revised Psychopathy Checklist*. Toronto, Canada: Multi-Health Systems; 2003.
- Harin V, Iennaco J, Olsen D. A review of ecological factors affecting inpatient psychiatric unit violence: Implications for relational and unit cultural improvements. *Issues in Mental Health Nursing*. 2009; 30:214–226. [PubMed: 19363726]

- Hart, SD.; Cox, DN.; Hare, RD. The Hare Psychopathy Checklist: Screening Version. Toronto, Canada: Multi Health Systems; 1995.
- Hart, SD.; Webster, CD.; Douglas, KS. Risk management using the HCR-20: A general overview of focusing on historical factors. In: Douglas, KS.; Webster, CD.; Hart, SD.; Eaves, D.; Ogloff, JRP., editors. HCR-20 violence risk management companion guide. Burnaby, Canada/Tampa, FL: Simon Fraser University, Mental Health, Law & Policy Institute/University of South Florida, Department of Mental Health Law & Policy; 2001. p. 27-40.
- Heilbrun, K.; Yasuhara, K.; Shah, S. Violence risk assessment tools: Overview and critical analysis. In: Otto, RK.; Douglas, KS., editors. Handbook of violence risk assessment. New York: Routledge/Taylor & Francis Group; 2010. p. 1-17.
- Higgins N, Watts D, Bindman J, Slade M, Thornicroft G. Assessing violence in general adult psychiatry. *Psychiatric Bulletin*. 2005; 29:131–133.
- Hilton NZ, Harris GT, Rice ME. Sixty-six years of research on the clinical versus actuarial prediction of violence. *The Counseling Psychologist*. 2006; 34:400–409.
- Johnstone L, Cooke DJ. PRISM: A promising paradigm for assessing and managing institutional violence: Findings from a multiple case study analysis of five Scottish prisons. *International Journal of Forensic Mental Health*. 2010; 9:180–191.
- Khiroya R, Weaver T, Maden T. Use and perceived utility of structured violence risk assessments in English medium secure forensic units. *Psychiatric Bulletin*. 2009; 33:129–132.
- Kooyman I, Dean K, Harvey S, Walsh E. Outcomes of public concern in schizophrenia. *British Journal of Psychiatry*. 2007; 191:s29–s36.
- Kroppan E, Nasset MB, Nonstad K, Pedersen TW, Almvik R, Palmstierna T. Implementation of the Short-Term Assessment of Risk and Treatability (START) in a forensic high secure unit. *International Journal of Forensic Mental Health*. 2011; 10:7–12.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]
- Lussier P, Verdun-Jones S, Deslauriers-Varin N, Nicholls TL, Brink J. Chronic violent patients in an inpatient psychiatric hospital: Prevalence, description, and identification. *Criminal Justice and Behavior*. 2010; 37:5–28.
- McDermott BE, Edens JF, Quanbeck CD, Busse D, Scott CL. Examining the role of static and dynamic risk factors in the prediction of inpatient violence: Variable- and person-focused analyses. *Law and Human Behavior*. 2008; 32:325–338. [PubMed: 17597388]
- Monahan, J.; Steadman, HJ.; Appelbaum, PS.; Grisso, T.; Mulvey, EP.; Roth, LH., et al. Classification of Violence Risk (COVR). Lutz, FL: Psychological Assessment Resources; 2005.
- Monahan, J.; Steadman, HJ.; Silver, E.; Appelbaum, P.; Robbins, P.; Mulvey, E., et al. Rethinking risk assessment: The Mac Arthur study of mental disorder and violence. New York: Oxford University Press; 2001.
- Nicholls TL, Brink J, Desmarais SL, Webster CD, Martin M-L. The Short-Term Assessment of Risk and Treatability (START): A prospective validation study in a forensic psychiatric sample. *Assessment*. 2006; 13:313–327. [PubMed: 16880282]
- Nicholls TL, Brink J, Greaves C, Lussier P, Verdun-Jones S. Forensic psychiatric inpatients and aggression: An exploration of incidence, prevalence, severity, and interventions by patient gender. *International Journal of Law and Psychiatry*. 2009; 32:23–30.
- Nicholls TL, Ogloff JRP, Douglas KS. Assessing risk for violence among female and male civil psychiatric patients: The HCR-20, PCL:SV, and McNiel & Binder's VSC. *Behavioral Sciences and the Law*. 2004; 22:127–158. [PubMed: 14963884]
- Nicholls, TL.; Petersen, K.; Brink, J. Field reliability of the START: The relationship between treatment team assessments and diverse patient safety events; Paper presented at the American Psychology-Law Society conference; 2011, March; Miami, FL.
- Nicholls TL, Petersen K, Brink J, Webster C. A clinical risk profile of forensic psychiatric patients: Treatment team STARTs in a Canadian service. *International Journal of Forensic Mental Health*. 2011; 10:187–199.
- Nonstad K, Nasset MB, Kroppan E, Pedersen TW, Nøttestad JA, Almvik R, Palmstierna T. Predictive validity and other psychometric properties of the Short-Term Assessment of Risk and Treatability

- (START) in a Norwegian high secure hospital. *International Journal of Forensic Mental Health Services*. 2010; 9:294–299.
- Otto, RK.; Douglas, KS., editors. *Handbook of violence risk assessment*. New York: Routledge; 2010.
- Petersen, KL.; Douglas, KS.; Nicholls, TL. Gender differences in the psychometric properties of the Short-Term Assessment of Risk and Treatability (START) in an acute civil psychiatric sample; Paper presented at the American Psychology-Law Society conference; 2011, March; Miami, FL.
- O'Shaughnessy RJ. AAPL practice guideline for the forensic psychiatric evaluation of competence to stand trial: A Canadian legal perspective. *Journal of the American Academy of Psychiatry and the Law*. 2007; 35:505–508. [PubMed: 18086743]
- Quinsey, VL.; Harris, GT.; Rice, ME.; Cormier, CA. *Violent offenders: Appraising and managing risk*. 2nd ed.. Washington, DC: American Psychological Association; 2006.
- Rice ME, Harris GT, Quinsey VL. The appraisal of violence risk. *Current Opinion in Psychiatry*. 2002; 15:589–593.
- Rogers R. The uncritical acceptance of risk assessment in forensic practice. *Law and Human Behavior*. 2000; 24:595–605. [PubMed: 11026213]
- Rutter M. Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry*. 1987; 5:316–331. [PubMed: 3303954]
- Ryba NL. The other side of the equation: Considering risk state and protective factors in violence risk assessment. *Journal of Forensic Psychology Practice*. 2008; 8:413–423.
- Saleebey, D. *The strengths perspective in social work practice*. 2nd ed.. New York: Addison-Wesley; 1996.
- Scott PD. Assessing dangerousness in criminals. *British Journal of Psychiatry*. 1977; 131:127–142. [PubMed: 912214]
- Simourd DJ. Use of dynamic risk/need assessment instruments among long-term incarcerated offenders. *Criminal Justice and Behavior*. 2004; 31:306–323.
- Singh JP, Grann M, Fazel S. A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*. 2011; 31:499–513. [PubMed: 21255891]
- Skeem JL, Monahan J. Current directions in violence risk assessment. *Current Directions in Psychological Science*. 2011; 20:38–42.
- Strand S, Belfrage H, Fransson G, Levander S. Clinical and risk management factors in risk prediction of mentally disordered offenders more important than historical data? A retrospective study of 40 mentally disordered offenders assessed with the HCR-20 violence risk assessment scheme. *Legal and Criminological Psychology*. 1999; 4:67–76.
- Verdun-Jones, S.; Brink, J.; Lussier, P.; Nicholls, TL. *Aggression and violence at the B.C. Forensic Psychiatric Hospital: Description, prediction, management and legal issues*. Vancouver, Canada: Forensic Psychiatric Services Commission; 2006.
- Viljoen, S.; Launeanu, M.; Hendry, M.; Nicholls, TL.; Brink, J. Comparing ratings of subject matter experts and clinicians on the START in a forensic clinic; Poster presented at the American Psychology-Law Society conference; 2011, March; Miami, FL.
- Vogel, V. de; Ruiters, C. de Differences between clinicians and researchers in assessing risk of violence in forensic psychiatric patients. *The Journal of Forensic Psychiatry & Psychology*. 2004; 15:145–164.
- Vogel, V. de; Ruiters, C. de Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime & Law*. 2006; 12:321–336.
- Vogel, V. de; Ruiters, C. de; Bouman, Y.; Vries Robbé, M. de *SAPROF. Guidelines for the assessment of protective factors for violence risk (English version)*. Utrecht, The Netherlands: Forum Educatief; 2009.
- Webster, CD.; Douglas, KS.; Eaves, D.; Hart, SD. *HCR-20: Assessing risk for violence (Version 2)*. Vancouver, Canada: Mental Health, Law, & Policy Institute, Simon Fraser University; 1997.
- Webster, CD.; Martin, ML.; Brink, J.; Nicholls, TL.; Desmarais, SL. *Manual for the Short-Term Assessment of Risk and Treatability (START) (Version 1.1)*. Coquitlam, Canada: British Columbia Mental Health & Addiction Services; 2009.

- Webster, CD.; Martin, ML.; Brink, J.; Nicholls, TL.; Middleton, C. Manual for the Short Term Assessment of Risk and Treatability (START) (Version 1.0, Consultation Edition). Hamilton, Canada: St. Joseph's Healthcare Hamilton, Ontario, Canada; Coquitlam, Canada: Forensic Psychiatric Services Commission; 2004.
- Webster CD, Nicholls TL, Martin ML, Desmarais SL, Brink J. Short-Term Assessment of Risk and Treatability (START): The case for a new violence risk structured professional judgment scheme. *Behavioral Sciences & the Law*. 2006; 24:747–766. [PubMed: 17171764]
- Wilson CM, Desmarais SL, Nicholls TL, Brink J. The role of client strengths in assessments of violence risk using the Short-Term Assessment of Risk and Treatability (START). *International Journal of Forensic Mental Health*. 2010; 9:282–293.
- Wilson CM, Desmarais SL, Nicholls TL, Hart SD, Brink J. Predictive validity of dynamic factors: Assessing violence risk in forensic psychiatric inpatients. (under review).
- Wong S. Is Hare's Psychopathy Checklist reliable without the interview? *Psychological Reports*. 1988; 62:931–934. [PubMed: 3406310]
- Yudofsky SC, Silver JM, Jackson W, Endicott J, Williams DW. The Overt Aggression Scale for the objective rating of verbal and physical aggression. *American Journal of Psychiatry*. 1986; 143:35–39. [PubMed: 3942284]

Appendix. START Summary Sheet (Version 1.1)

BC Mental Health &
Addiction Services
An Agency of the Provincial Health Services Authority

St. Joseph's
Healthcare Hamilton

Name: _____
Last First Middle

Record #: _____

Male ☐ Female ☐ D.O.B.: _____
yy / mm / dd

START Summary Sheet ©

Diagnoses: DSM-IV^{TR} ☐ ICD-10 ☐ 1 _____ 2 _____
3 _____ 4 _____ 5 _____

STATUS: ☐ HOSPITAL ☐ COMMUNITY ☐ CORRECTIONS
Status: _____

PURPOSE: ☐ REFERRAL ☐ ADMISSION ☐ REVIEW ☐ OTHER
Specify: _____

START Time Frame: _____
days / weeks / months

Key Item	Strengths			START Items	Vulnerabilities			Critical Item	SIGNATURE RISK SIGNS					
	2	1	0		0	1	2							
☐	☐	☐	☐	1. Social Skills	☐	☐	☐	☐						
☐	☐	☐	☐	2. Relationships (TA:Y/N)*	☐	☐	☐	☐						
☐	☐	☐	☐	3. Occupational	☐	☐	☐	☐						
☐	☐	☐	☐	4. Recreational	☐	☐	☐	☐						
☐	☐	☐	☐	5. Self-Care	☐	☐	☐	☐	Hx*	Risks	T.H.R.E.A.T.	Low	Mod	High
☐	☐	☐	☐	6. Mental State	☐	☐	☐	☐	Violence	No ☐	Yes ☐	☐	☐	☐
☐	☐	☐	☐	7. Emotional State	☐	☐	☐	☐	Self-Harm	No ☐	Yes ☐	☐	☐	☐
☐	☐	☐	☐	8. Substance Use	☐	☐	☐	☐	Suicide	No ☐	Yes ☐	☐	☐	☐
☐	☐	☐	☐	9. Impulse Control	☐	☐	☐	☐	Unauthorized Leave		☐	☐	☐	
☐	☐	☐	☐	10. External Triggers	☐	☐	☐	☐	Substance Abuse		☐	☐	☐	
☐	☐	☐	☐	11. Social Support (PPS:Y/N) [‡]	☐	☐	☐	☐	Self-Neglect		☐	☐	☐	
☐	☐	☐	☐	12. Material Resources	☐	☐	☐	☐	Being Victimized		☐	☐	☐	
☐	☐	☐	☐	13. Attitudes	☐	☐	☐	☐	Case Specific Risk:		☐	☐	☐	
☐	☐	☐	☐	14. Med. Adherence (N/A ☐ [‡])	☐	☐	☐	☐	CURRENT MANAGEMENT MEASURES					
☐	☐	☐	☐	15. Rule Adherence	☐	☐	☐	☐						
☐	☐	☐	☐	16. Conduct	☐	☐	☐	☐						
☐	☐	☐	☐	17. Insight	☐	☐	☐	☐						
☐	☐	☐	☐	18. Plans	☐	☐	☐	☐						
☐	☐	☐	☐	19. Coping	☐	☐	☐	☐						
☐	☐	☐	☐	20. Treatability	☐	☐	☐	☐	Current Management Plan:					
☐	☐	☐	☐	21. Case Specific Item:	☐	☐	☐	☐						
☐	☐	☐	☐	22. Case Specific Item:	☐	☐	☐	☐						

Health Concerns/Medical Tests:

Risk Formulation: what factors/predict-explain/which person/will carry out/what act/when?

COMPLETED BY: _____ DATE: _____

TA - Therapeutic Alliance [‡]PPS - Positive Peer Support [‡]N/A - Not Applicable ^{}Hx - Historical

Version 1.1 © 2009

Table 1

Nature and Severity of Aggression

Aggressive Behaviors	Number of Participants who Engaged in Aggression	Frequency of Incidents among Participants who Engaged in Some Form of Aggression	
	<i>n</i> (%)	<i>M</i> (<i>SD</i>)	Range
Any Aggression	65 (54.2)	17.25 (34.64)	1 – 196
Verbal Aggression (<i>n</i> = 119)	63 (52.5)	13.47 (26.28)	0–142
Level 1	56 (46.7)	6.08 (10.22)	0–49
Level 2	30 (25.0)	1.23 (2.57)	0–17
Level 3	46 (38.3)	5.44 (13.14)	0–75
Level 4	20 (16.7)	1.27 (2.82)	0–17
Physical Aggression - Objects	20 (16.7)	2.66 (6.71)	0–33
Level 1	16 (13.3)	1.71 (4.61)	0–29
Level 2	14 (11.7)	0.86 (2.58)	0–17
Level 3	4 (3.3)	0.14 (0.69)	0–5
Level 4	2 (1.7)	0.06 (0.39)	0–3
Physical Aggression - Others	27 (22.5)	1.36 (3.05)	0–21
Level 1	16 (13.3)	0.47 (1.10)	0–5
Level 2 (<i>n</i> = 118)	16 (13.3)	0.53 (1.30)	0–8
Level 3	10 (8.3)	0.30 (1.19)	0–9
Level 4	1 (0.8)	0.03 (0.25)	0–2

Note. *N* = 120 unless otherwise specified. Measured using a revised version of the OAS.

* Although the OAS includes verbal aggression to self in Level 4, verbal aggression in this table refers only to aggression to others.

Table 2

Descriptive Statistics for START, HCR-20 and PCL:SV Scores

Assessment Approach	Descriptive Statistics		
	<i>M</i>	<i>SD</i>	Range
START			
Strength Total Score	18.46	8.08	4 – 38
Vulnerability Total Score	16.82	8.07	0 – 33
HCR-20			
Historical Subscale Score	13.82	3.41	4 – 20
Clinical Subscale Score	4.81	2.51	0 – 10
Risk Management Subscale Score	6.19	2.36	0 – 10
Total Score	24.90	6.59	7 – 38
PCL:SV			
Total Score	11.72	4.20	0 – 23

Note. *N* = 120. Possible range of START Strength and Vulnerability total scores is 0 – 40, where higher scores indicate greater strength and vulnerability, respectively. Possible range of HCR-20 Historical subscale scores is 0 – 20, Clinical and Risk Management subscale scores is 0 – 10, and total scores is 0 – 40. Higher scores indicate greater risk. Possible range for PCL:SV total scores is 0 – 24.

Table 3

START Item Score Summaries

Item		Descriptives		Frequency of Endorsement (%)			<i>r</i>
		<i>M</i>	<i>SD</i>	Minimally Present (0)	Moderately Present (1)	Maximally Present (2)	
1. Social Skills	Strength	0.97	0.65	22.5	58.3	19.2	-.59***
	Vulnerability	0.94	0.71	28.3	49.2	22.5	
2. Relationships	Strength	0.78	0.64	33.3	55.0	11.7	-.41***
	Vulnerability	0.69	0.67	42.5	45.8	11.7	
3. Occupational	Strength	0.99	0.70	25.0	50.8	24.2	-.67***
	Vulnerability	0.62	0.70	50.8	36.7	12.5	
4. Recreational	Strength	1.20	0.66	13.3	53.3	33.3	-.56***
	Vulnerability	0.53	0.56	50.0	46.7	03.3	
5. Self-Care (<i>n</i> = 119)	Strength	1.24	0.64	10.9	53.8	35.3	-.70***
	Vulnerability	0.54	0.66	55.5	35.3	09.2	
6. Mental State	Strength	0.99	0.69	24.2	52.5	23.3	-.65***
	Vulnerability	0.97	0.74	29.2	45.0	25.8	
7. Emotional State	Strength	1.16	0.61	11.7	60.8	27.5	-.55***
	Vulnerability	0.79	0.71	37.5	45.8	16.7	
8. Substance Use	Strength	1.05	0.82	30.8	33.3	35.8	-.84***
	Vulnerability	0.81	0.84	46.7	25.8	27.5	
9. Impulse Control	Strength	0.94	0.73	29.2	47.5	23.3	-.80***
	Vulnerability	0.83	0.81	42.5	31.7	25.8	
10. External Triggers	Strength	0.80	0.74	39.2	41.7	19.2	-.62***
	Vulnerability	0.93	0.71	29.2	49.2	21.7	
11. Social Support	Strength	0.78	0.69	36.7	48.3	15.0	-.80***
	Vulnerability	1.18	0.78	22.5	36.7	40.8	
12. Material Resources	Strength	0.93	0.51	16.7	74.2	09.2	-.52***
	Vulnerability	0.89	0.55	20.8	69.2	10.0	
13. Attitudes (<i>n</i> = 119)	Strength	0.94	0.72	28.6	48.7	22.7	-.71***
	Vulnerability	0.71	0.75	46.2	36.1	17.6	
14. Medication Adherence (<i>n</i> = 118)	Strength	1.19	0.64	12.7	55.1	32.2	-.60***

Item	Descriptives		Frequency of Endorsement (%)			<i>r</i>
	<i>M</i>	<i>SD</i>	Minimally Present (0)	Moderately Present (1)	Maximally Present (2)	
15. Rule Adherence	Vulnerability	0.45	0.61	61.0	33.1	05.9
	Strength	1.13	0.74	21.7	44.2	34.2
	Vulnerability	0.73	0.78	47.5	32.5	20.0
16. Conduct	Strength	1.00	0.75	27.5	45.0	27.5
	Vulnerability	0.68	0.78	50.8	30.0	19.2
17. Insight	Strength	0.50	0.62	56.7	36.7	6.7
	Vulnerability	1.49	0.65	08.3	34.2	57.5
18. Plans	Strength	0.55	0.63	52.5	40.0	07.5
	Vulnerability	1.16	0.79	24.2	35.8	40.0
19. Coping	Strength	0.64	0.62	43.3	49.2	7.5
	Vulnerability	0.99	0.64	20.8	59.2	20.0
20. Treatability	Strength	0.71	0.64	39.2	50.8	10.0
	Vulnerability	0.90	0.72	30.8	48.3	20.8

Note. *N* = 120 unless otherwise specified. *r* = correlation between Strength and Vulnerability scores.

p < .001.

Table 4

Correlations between START, HCR-20 and PCL:SV Scores

Assessment Approach	1	2	3	4	5	6	7
START							
1. Strength Total Score	--						
2. Vulnerability Total Score	-.87**	--					
HCR-20							
3. Historical Subscale Score	-.43***	.46***	--				
4. Clinical Subscale Score	-.72***	.81***	.33***	--			
5. Risk Management Subscale Score	-.76***	.77***	.37***	.64***	--		
6. Total Score	-.77***	.83***	.79***	.78***	.80***	--	
PCL:SV							
7. Total Score	-.13	.21*	.34***	.10	.11	.25**	--

Note. N= 120 unless otherwise specified. Higher Strength total scores indicate greater strength and higher Vulnerability total scores, Historical, Clinical, and Risk Management subscale scores, and HCR-20 total scores indicate greater risk. PCL:SV total scores indicate greater endorsement of psychopathic traits.

* p < .05.
** p < .01.
*** p < .001.

Table 5

Distribution of START and HCR-20 Violence Risk Estimates

HCR-20	START		
	Low <i>n</i> (%)	Moderate <i>n</i> (%)	High <i>n</i> (%)
Low	42 (35.3)	5 (4.2)	0 (0)
Moderate	4 (3.4)	21 (17.6)	5 (4.2)
High	0 (0)	4 (3.4)	33 (27.7)

Note. *N* = 119.

Table 6

Frequency (and Percentage) of Participants Who Engaged in Aggression as a function of START Violence Risk Estimates

Aggressive Behaviors	START Violence Risk Estimate		
	Low <i>n</i> = 46	Moderate <i>n</i> = 40	High <i>n</i> = 33
Any Aggression	9 (19.6%)	27 (67.5%)	28 (84.8%)
Verbal Aggression	10 (21.7%)	25 (62.5%)	28 (84.8%)
Physical Aggression – Objects	0 (0%)	5 (12.5%)	14 (42.4%)
Physical Aggression – Others	1 (2.2%)	6 (15.0%)	20 (60.6%)

Note. *N* = 119. Percentages are calculated within the rating of violence risk.

Table 7
Validity of START, HCR-20 and PCL:SV Assessments in Predicting Inpatient Aggression

Assessment	Inpatient Aggression					
	Any Aggression		Verbal		Physical - Objects	
	AUC (SE)	95% CI	AUC (SE)	95% CI	AUC (SE)	95% CI
START						
Strength Total Score	.76 (.04) ***	.68-.85	.75 (.05) ***	.66-.84	.77 (.06) ***	.65-.89
Vulnerability Total Score	.79 (.04) ***	.71-.87	.79 (.04) ***	.70-.87	.80 (.05) ***	.71-.89
Violence Risk Estimate (<i>n</i> = 119)	.80 (.04) ***	.72-.88	.78 (.04) ***	.69-.86	.84 (.04) ***	.76-.92
HCR-20						
Historical Subscale Score	.73 (.05) ***	.64-.82	.71 (.05) ***	.62-.80	.66 (.06) ***	.54-.78
Clinical Subscale Score	.74 (.05) ***	.65-.83	.74 (.05) ***	.65-.83	.78 (.05) ***	.68-.89
Risk Management Subscale Score	.77 (.04) ***	.69-.86	.77 (.05) ***	.69-.86	.77 (.05) ***	.66-.87
Total Score	.80 (.04) ***	.72-.88	.80 (.04) ***	.72-.88	.79 (.05) ***	.69-.89
Violence Risk Estimate	.79 (.04) ***	.71-.87	.74 (.06) ***	.62-.86	.70 (.10) ***	.51-.89
PCL:SV						
Total Score	.75 (.05) ***	.67-.84	.74 (.05) ***	.65-.83	.63 (.07) ***	.55-.76

Note. *N* = 120 unless otherwise specified. Values are Areas under the Curve (AUC) for Receiver Operating Characteristic (ROC) curves. SE = Standard Error. CI = Confidence Interval. Strength total scores are reverse coded such that lower scores indicate greater strength.

* *p* < .05.

** *p* < .01.

*** *p* < .001.

Table 8
Logistic Regression Analyses Testing the Incremental Validity of START Assessments over HCR-20 Historical Subscale Scores

	Any Aggression			Verbal Aggression			Physical – Objects			Physical – Others		
	β	Wald	Odds Ratio	β	Wald	Odds Ratio	β	Wald	Odds Ratio	β	Wald	Odds Ratio
Step 1	Model fit $\chi^2(1) = 17.93^{***}$ Model fit $\chi^2(1) = 15.36^{***}$ Model fit $\chi^2(1) = 5.29^*$ Model fit $\chi^2(1) = 10.68^{***}$											
Historical Subscale Score	0.25 ^{***}	15.02	1.29	0.23 ^{***}	13.17	1.26	0.19 [*]	4.63	1.21	0.24 ^{**}	8.87	1.27
Step 2	Model fit $\chi^2(3) = 37.26^{***}$ Model fit $\chi^2(3) = 34.90^{***}$ Model fit $\chi^2(3) = 18.81^{***}$ Model fit $\chi^2(3) = 28.21^{***}$ $\Delta\chi^2(2) = 19.33^{***}$ $\Delta\chi^2(2) = 19.55^{***}$ $\Delta\chi^2(2) = 13.51^{***}$ $\Delta\chi^2(2) = 17.53^{***}$											
Historical Subscale Score	0.16 [*]	4.03	1.17	0.13	3.13	1.14	0.07	0.46	1.07	0.17	3.02	1.18
Strength Total Score	0.02	0.08	1.02	0.00	0.01	1.00	0.05	0.47	1.05	0.15 [*]	4.67	1.16
Vulnerability Total Score	0.12 [*]	4.05	1.13	0.13 [*]	5.27	1.14	0.10	2.22	1.11	0.01	0.04	1.01
Step 3	Model fit $\chi^2(4) = 45.53^{***}$ Model fit $\chi^2(4) = 41.31^{***}$ Model fit $\chi^2(4) = 29.04^{***}$ Model fit $\chi^2(4) = 44.36^{***}$ $\Delta\chi^2(1) = 8.27^{***}$ $\Delta\chi^2(1) = 6.41^{**}$ $\Delta\chi^2(1) = 10.23^{***}$ $\Delta\chi^2(1) = 16.16^{***}$											
Historical Subscale Score	0.11	2.02	1.12	0.09	1.39	1.10	0.01	0.01	0.99	0.08	0.60	1.09
Strength Total Score	0.02	0.11	1.02	0.01	0.01	1.01	0.03	0.14	1.03	0.14 [*]	3.71	1.15
Vulnerability Total Score	0.05	0.79	1.06	0.07	1.46	1.08	0.04	0.23	1.04	0.08	1.33	0.93
Violence Risk Estimate	1.07 ^{**}	7.61	2.91	0.92 [*]	6.04	2.50	1.68 ^{**}	8.11	5.37	1.85 ^{***}	12.52	6.39

Note: Strength total scores are reverse coded such that lower scores indicate greater strength.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 9
Logistic Regression Analyses Testing the Incremental Validity of START Assessments over PCL:SV Total Scores

	Any Aggression			Verbal Aggression			Physical – Objects			Physical – Others		
	β	Wald	Odds Ratio	β	Wald	Odds Ratio	β	Wald	Odds Ratio	β	Wald	Odds Ratio
Step 1	Model fit $\chi^2(1) = 21.72^{***}$ Model fit $\chi^2(1) = 19.13^{***}$ Model fit $\chi^2(1) = 3.69$ Model fit $\chi^2(1) = 15.67^{***}$											
PCL:SV Total Score	0.24 ^{***}	16.42	1.27	0.22 ^{**}	14.94	1.25	0.12	3.50	1.13	0.23 ^{***}	12.55	1.26
Step 2	Model fit $\chi^2(3) = 36.74^{***}$ Model fit $\chi^2(3) = 34.62^{***}$ Model fit $\chi^2(3) = 18.44^{***}$ Model fit $\chi^2(3) = 29.14^{***}$ $\Delta\chi^2(2) = 15.02^{***}$ $\Delta\chi^2(2) = 15.41^{***}$ $\Delta\chi^2(2) = 14.75^{***}$ $\Delta\chi^2(2) = 13.48^{***}$											
PCL:SV Total Score	0.13	3.71	1.14	0.11	2.79	1.11	0.03	0.10	0.97	0.15 [*]	3.94	1.16
Strength Total Score	0.01	0.07	1.01	0.00	0.00	1.00	0.05	0.45	1.05	0.14 [*]	4.33	1.15
Vulnerability Total Score	0.11 [*]	3.91	1.12	0.12 [*]	4.70	1.13	0.12	3.06	1.13	0.00	0.00	1.00
Step 3	Model fit $\chi^2(4) = 44.40^{***}$ Model fit $\chi^2(4) = 50.52^{***}$ Model fit $\chi^2(4) = 31.68^{***}$ Model fit $\chi^2(4) = 21.72^{***}$ $\Delta\chi^2(1) = 7.65^{***}$ $\Delta\chi^2(1) = 5.90^{*}$ $\Delta\chi^2(1) = 13.23^{***}$ $\Delta\chi^2(1) = 14.83^{***}$											
PCL:SV Total Score	0.07	0.89	1.07	0.06	0.60	1.06	0.15	2.24	0.86	0.04	0.22	1.04
Strength Total Score	0.02	0.15	1.02	0.01	0.02	1.01	0.03	0.14	1.03	0.14	3.62	1.15
Vulnerability Total Score	0.05	0.71	1.05	0.07	1.38	1.07	0.05	0.42	1.05	0.08	1.23	0.93
Violence Risk Estimate	1.06 ^{**}	7.06	2.89	0.91 [*]	5.57	2.49	2.08 ^{**}	9.77	7.98	1.85 ^{***}	11.53	6.38

Note: Strength total scores are reverse coded such that lower scores indicate greater strength.

* $p < .05$.

** $p < .01$.

*** $p < .001$.