

THE PREDICTIVE VALIDITY OF CLINICAL RATINGS OF THE SHORT-TERM ASSESSMENT OF RISK AND TREATABILITY (START)

Erika Braithwaite, Yanick Charette

Université de Montréal and Douglas Mental Health University Institute, Québec,
Canada

Anne G. Crocker

Douglas Mental Health University Institute and McGill University, Québec, Canada
Andrea Reyes

Douglas Mental Health University Institute, Québec, Canada

Paper published in International Journal of Forensic Mental Health, 2011, 9(4)

ABSTRACT

With the increased need to assess and manage risk in inpatient settings, the Short-Term Assessment of Risk and Treatability (START) was implemented on a civil psychiatric unit. The goal of the present study was to examine the tool's predictive validity when completed by clinical teams as part of routine practice. Data were collected for 34 patients hospitalized for a minimum of 30 days prior to and after a START evaluation. Several challenging behaviors, such as aggression towards others, self-harm, and substance abuse were assessed using the START Outcomes Scale (Nicholls et al., 2007). Results from multilevel logistic regression and Receiver Operating Characteristics analyses lend partial support for the predictive validity of the START. A limited set of START items combined was significantly better at predicting the challenging behaviors than the original total Strength and Vulnerability scales. Results are discussed in terms of the clinical use of risk assessment.

KEYWORDS: START, risk assessment, risk management, protective factors, civil psychiatric setting

Aggressive behavior displayed by individuals with a mental illness in inpatient settings can pose significant challenges to both staff and patients. Rates of aggressive behavior vary as a function of the definition, operationalization of the behaviors and the setting (Nijman, 1999). When measuring a broad range of problematic behaviors, rates from 13% (Barlow, Grenyer, & Ilkiw-Lavalle, 2000; Duxbury, 2002) to 60% (Nicholls, Brink, Desmarais, Webster, & Martin, 2006) have been found. However, mere

numbers do not adequately portray the complexity of inpatient aggression. Almost 100% of mental health care workers surveyed in a Swedish study had been a victim of aggression at one point or another in their careers (Menckel & Viitasara, 2002). Frontline workers in psychiatric settings are continuously making decisions relating to admissions, level of security and privileges, and discharges (McNeil, Gregory, Lam, Binder, & Sullivan, 2003). However, most structured risk assessment tools are designed

to assess the long-term risk of violence to others (McNiels et al., 2003; St'ubner, Groß, & Nedopil, 2006). Tools providing assessment of long-term risk are of limited use in the day-to-day management of risk with individuals with a severe mental health problem in inpatient psychiatric care.

The domain of violence risk assessment has evolved tremendously over the past 20 years. There has been a distinct shift from the prediction of dangerousness to the management of risk (e.g., Doyle & Dolan, 2002). One of the most notable developments has been risk assessment instruments using the structured professional judgment (SPJ) approach. The SPJ approach is centered on assessment tools made up of empirically validated risk factors that aim to help mental health professionals structure their clinical assessments of patients (Doyle & Dolan, 2008). Clinicians use the risk factors to inform risk management decisions and plan treatment strategies.

Protective Factors in Risk Assessment

While the study of risk factors for aggression and violence has blossomed over the past years, the almost complete exclusion of an examination of protective factors is remarkable (Sheldrick, 1999). The traditional model of risk assessment focuses almost entirely on the prediction of aggressive and violent acts, with the use of risk factors shown to increase the likelihood of their occurrence. Conversely, protective factors are variables that reduce the effect of risk factors or influence the outcome independently, in this case, the likelihood of aggression and violence.

Some efforts are being made to develop risk assessment instruments that include strengths or protective factors in the prediction of risk for aggression and violence. The Inventory of Offender Risks, Needs, and Strengths (IORNS; Miller, 2006) is a self-report measure designed to predict aggression and long-term recidivism. The IORNS contains 130 yes/no items that assess static risks, dynamic needs,

and protective strengths. The protective strength variables relate to cognitive/behavioral regulation, anger regulation, and education/training. One study to date has evaluated the IORNS's accuracy in predicting general recidivism (Miller, 2006). In this study of offenders released to a halfway house, recidivism was defined as any violation of the halfway house rules that merited the offender being referred back to the prison for assessment. Of the 162 male offenders followed, 22% committed at least one infraction during the 15-month follow-up period. The majority of incidents were of mild severity. Those who recidivated on two or more occasions had higher scores on the overall risk index and the dynamic needs index and lower scores on the protective strengths index than those who had recidivated once or not at all.

The Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2006) is a tool designed to predict violence among adolescents. It is based on the structured professional judgment approach. The SAVRY assesses static (historical), socio-contextual factors and individual (clinical) factors and items are scored as either present, possibly or mildly present or absent. It also contains six protective factors rated as either present or absent. The protective factors include community involvement, social support and personality traits. Results from one file based prospective study (Lodewijks, Doreleijers, de Ruiter, & Borum, 2008) suggest that the SAVRY protective factors add to the incremental validity of the total risk score in the prediction of general recidivism.

One of the most recent and promising additions to the study of protective factors in violence risk is the Structured Assessment of PROtective Factors for violence risk (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbé, 2009). The SAPROF is based on the SPJ approach, and is made up entirely of protective factors that relate to internal, motivational and external factors. It was designed to be used in combination with other structured guidelines (de Vogel et al., 2009) to provide a more

balanced perspective of patients. While still in its initial stages of development, preliminary research suggests that the SAPROF has excellent interrater reliability and good predictive validity (de Vogel et al., 2009).

Another recent and promising instrument, the Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Middleton, 2004) is meant to be used through an SPJ approach and the subject of the current research. The START is designed to aid in the management of various types of risk in inpatient settings. It is made up of 20 dynamic items that are rated on both strength and vulnerability scales (Webster et al., 2004). Three critical advantages of the START are: inclusion of protective factors, short term time predictions, and prediction of multiple types of adverse outcomes (violence, self-harm, suicidality, unauthorized leave, substance abuse, self-neglect, and victimization).

The IORNS, SAVRY, SAPROF, and START indicate that it is possible to incorporate protective factors into risk assessment instruments. However, the extant literature still lacks a conceptual framework explaining why and how protective factors interact with or affect risk. Studies that have included protective factors have noted that determining whether a factor is considered a risk or strength depends largely on the researcher's intuition (e.g., Gagliardi, Lovell, Peterson, & Jemelka, 2004). How protective factors are defined and operationalized has not been standardized, leading researchers to define these variables in different ways. This in turn makes it more difficult to compare study results and draw conclusions about the usefulness of including protective factors in risk assessment schemes. Despite these limitations, clinicians have reported that protective factors were as important to clinical practice as risk factors (Stübner et al., 2006).

There are different ways in which strengths and risks can be conceptualized. The two may exist on

different ends of the same spectrum (Brook, Whiteman, Gordon, & Cohen, 1989; Rutter, 1987), suggesting that an individual may be at either the high or low end, but not both simultaneously. Conversely, strengths could be qualitatively different from risks and an individual may score both high and low in a particular area. Newcomb and Felix-Ortiz's views (1992) are in line with this second notion and concede that "the assumption that absence of risk is equivalent to protection lacks validation and overlooks potential differences between the risk and protective potency of specific factors" (p. 281). The START uses this framework to allow clinical team members to rate each item of the instrument on both the strength scale and vulnerability scale. An individual might have social support from one group of friends (strength) but also have a disinterested family (vulnerability). The protective factors in the START are called Strengths, and the risks are called Vulnerabilities. In a critical review of risk assessment, Rogers (2000) suggested that risk evaluations that do not incorporate protective factors are inherently inaccurate and unethical. One-sided assessments provide a biased portrait of offenders and perpetuate the perception that risk of violence and aggression is chronic and irreversible.

Short-term Assessment and Dynamic Risk

The conceptualization of risk as changing and dynamic is intuitive at the clinical level. However, a great deal of research in the field has focused on long-term (1 year and beyond) prediction of risk (Almvik, Woods, & Rasmussen, 2000). The importance of short-term prediction in inpatient settings is becoming increasingly important as the tendency is towards shorter hospitalization (Narrow, Regier, Rae, Manderscheid, & Locke, 1993). Risk assessment is no longer about making dichotomous predictions of violence, but rather the "ongoing, day-to-day decisions about management and treatment" (Steadman et al., 1993, p. 41).

The time periods intended to be captured by risk assessment instruments and how often dynamic risk

factors should be measured to understand their variability remains a challenge for researchers (Douglas & Skeem, 2005). Often, the developers of the instrument designate assessment tools as short-term. Studies have conceptualized short-term as being 24 hours (Almvik et al., 2000; Ogloff&Daffern, 2006), a few days (McNiel & Binder, 1995; McNiel et al., 2003), a few months (Douglas & Ogloff, 2003) up to one year (Nicholls et al., 2006). Despite these time span differences, all underline the importance of dynamic risk factors when predicting short-term risk.

The use of dynamic risk factors has proven to be especially crucial in inpatient settings. Traditionally, actuarial assessment tools that rely on historical factors have been able to accurately predict violence occurring in the community after discharge (Wang & Diamond, 1999). However, the need to predict, assess and manage these behaviors in institutional settings is undeniable. In fact, dynamic factors have been found to ‘override’ well-established historical factors in predicting the risk of short-term violence in institutional settings (McNiel et al., 2003). Others have replicated these findings. Grevatt and colleagues (2004) used the Historical and Clinical subscales of the HCR-20 (Webster, Douglas, Eaves, & Hart, 1997) to predict institutional violence in a secure psychiatric facility over a six-month period. The authors found that the clinical/dynamic factors were better at predicting repetitive violence ($AUC = .79$) than the historical/static factors ($AUC=.41$). Another study found that once patients are released in the community, and their psychiatric symptoms have presumably stabilized (McNiel et al., 2003), the Historical and Risk scales of the HCR-20 continue to predict long-term recidivism in previously civilly committed patients (Douglas, Ogloff, Nicholls, & Grant, 1999). These findings illustrate the importance of using dynamic factors when predicting short-term risk in psychiatric settings.

Instruments such as the START, which rely solely on dynamic risk factors, are useful in clinical settings because they can be used for a variety of contexts and outcomes. The developers of the START

(Webster et al., 2004) recommend that it be administered whenever there are signs of changes in vulnerabilities or strengths, upcoming changes in the individual’s legal status or when the treatment plan calls for an evaluation. Short-term prediction is useful because it takes into account the individual’s current mental state and (changing) environment. This in turn, enables clinicians to make appropriate changes to treatment plans according to the evolution of the patient.

Multiple Risk Outcomes

The study of violence and aggression among forensic and psychiatric clienteles focuses largely on physical aggression directed toward others (Ferris et al., 1997). However, mental health care workers deal with a much broader range of challenging behavior on a day-to-day basis (Crocker et al., *in press*). This reality is especially true given the increasing number of patients with forensic histories being seen in civil psychiatric settings (Crocker & Côté, 2009; Hodgins et al., 2007) Patients may pose a threat to staff members and other patients, but also to themselves (e.g., self-mutilation, suicidal behavior, self neglect) and pose risks for other types of challenging behaviors on a psychiatric unit (e.g., unauthorized leave, substance abuse). This also applies to behaviors that might appear less important for physical safety, but still pose significant management challenges and can be an obstacle to community-based service integration (e.g., verbal aggression). The notion that various types of risks exist, not simply that of physical aggression, needs further study (Towl, 2005).

Choe, Teplin, and Abrams (2008) conducted a review of empirical studies of perpetration of violence and violent victimization among individuals suffering from severe mental illnesses. The authors found that while both types of behaviors were common in inpatient and outpatient samples, there were no investigations of both types of behaviors within the same sample. Individuals with severe mental illness were 16 times more likely to be

victimized than to have offended violently. In a recent Canadian study (Joyal, Gendron, & Côté, 2008) that measured verbal, self-harm, physical, and aggression against objects, the authors found that while over half of the patients on the unit in a forensic psychiatric institution had engaged in at least one of the aforementioned behaviors, most incidents were mild in severity. More importantly, each incident required an average of 1.36 interventions per patient (e.g., isolation, seclusion or administration of medication). These studies and others (e.g., Nijman, Bowers, Oud, & Jansen, 2005) suggest that milder forms of aggression are common in inpatient psychiatric settings and warrant exploration.

Few studies examine a broad range of challenging behaviors, but even fewer attempt to predict them. Lodewijks, Doreleijers, De Ruiter and Borum (2008) recently used the SAVRY (described above) to predict various types of problematic behaviors in youths referred to a correctional treatment facility. The SAVRY was shown to have good predictive validity for physical aggression, aggression against objects, verbal threats and rule violation (AUCs ranging from .58 to .80). This shift to predicting several types of risks is an important step in the move towards an understanding that individuals can pose many different kinds of threats and tools that incorporate these many facets will be assets in many treatment and management settings.

In summary, the extant literature was missing a tool that incorporates protective factors, short-term risk assessment and the prediction of multiple risks. The START is still in its early stages of validation, but thus far, the results appear promising. Preliminary results have shown that the START has proved useful for nurses in a forensic setting (Doyle, Lewis, & Brisbane, 2008). Nicholls and colleagues (2006) tested the START in a Canadian forensic psychiatric hospital. Patient behavior was rated using a modified version of the Overt Aggression Scale (OAS; Yudofsky, Silver, Jackson, Enticott, & Williams, 1986). The OAS was modified to include the risk domains of the START as well as other behaviors

that pose management challenges on psychiatric inpatient units (Nicholls et al., 2007). The authors found that the START had good interrater reliability ($ICC=.87$) and internal consistency (Cronbach's alpha $\alpha = .87$). With respect to predictive validity, the START total scores had moderate statistically significant associations with the behaviors on the OAS (AUCs ranging from .31 to .92) occurring over a one year period. Crocker and colleagues (2008) implemented the START on a risk management civil psychiatric unit. The authors found that START vulnerability scores were predictive of physical aggression as well as aggression against objects in the 1, 6, and 12 months following the assessment and were not statistically predictive of the other challenging behaviors measured. Inversed START strength scores were also predictive of physical aggression against persons as well as aggression against objects at 1, 3, 6, and 12 months.

What is clear from the research with the START, as with many other risk assessment schemes, is that work is missing on the predictive validity of clinicians' ratings, and not simply those of trained research assistants or researchers. Few studies have used clinicians' actual risk assessments to conduct their analyses. Thus, START assessments completed by front line staff on a civil psychiatric unit (Crocker et al., in press; Crocker et al., 2008) were used to predict inpatient challenging behaviors. More specifically, the goal of the present study was to test (a) the relationship between the items of the strength and vulnerability scales in predicting the occurrence of a behavior within a 30-day time span, (b) the relationship between the strength and vulnerability scales and the occurrence of a behavior, (c) the relationship between the START risk estimates and the actual occurrence of a behavior within a 30-day time span, (d) to develop "optimized scales" to test the relationship between correlated items and the occurrence of a behavior.

METHOD

This longitudinal prospective study was conducted over two years. The START was implemented (see Crocker et al., in press, for more details) on the 16-bed Risk management and rehabilitation unit (RMRU) of a civil psychiatric hospital, the Douglas Institute of Montreal. Following implementation, the START has continued to be used on the RMRU.

Participants

A total of 84 patients were admitted onto the RMRU during the course of the study. The constant movement of patients on the unit through repeated discharges and admissions consequently created varying follow-up periods and numbers of START assessment per individual. A subsample of 34 patients who had been present continuously on the unit 30 days before and 30 days after a START assessment by the clinical team were included for the purpose of the current paper. A period of 60 days allowed not only for the opportunity of the staff to conduct a reliable START assessment, but it also permitted each participant with the same opportunity to engage in a behavior. These 34 patients received a total of 133 START assessments over the course of their stay on the unit. In order to verify how representative our subsample was for all patients on the unit, Chi-square and t -tests were conducted

comparing the current sample to those who were on the unit for shorter periods of time. Table 1 provides sociodemographic information for both the entire unit and the subsample ($n = 34$). Only the number of days on the unit was significantly different between the two groups, $t(116) = 8.20, p < .05, d = 1.82$.

Measures

Sociodemographic variables. Sociodemographic information was gathered through file review and consisted of gender, age, language, and civil status.

Psychopathological variables.

Psychopathological information included psychiatric diagnosis, and number of previous admissions and emergency visits to the Douglas Institute. This information was gathered through file review.

Challenging behaviors. Challenging behaviors were defined according to the START outcome categories and gathered through consultation of the notes kept by the unit's nurses. The behaviors were operationalized using the START-Outcome scale (SOS; Nicholls et al., 2007). This scale was constructed using a modified version of the Overt Aggression Scale (OAS; Yudofsky et al., 1986). Eleven types of challenging behaviors were gathered through review of the patient files and coded according to the SOS by trained research assistants. These eleven challenging behaviors were: physical

Table 1 : Sociodemographic information for the unit sample and subsample

	Unit ^a		Subsample ^b		χ^2	df
	n	%	n	%		
Male gender	34	68	27	79.4	1.33	1
Never married	40	80	30	88.2	0.99	1
Canadian (vs. other)	40	80	26	76.5	0.15	1
English (vs. other)	22	44	16	47.1	0.08	1
Diagnosis						
Schizophrenia	37	74	30	88.2	2.54	1
Mood disorder	7	14	4	11.8	0.09	1
Substance Abuse	12	24	9	26.5	0.07	1
Personality Disorder	16	32	7	20.6	1.33	1
Intellectual disability	5	10	6	17.6	1.04	1
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	
Age at implementation of START	40.04	13.38	37.91	11.72	0.75	82
Number of hospitalizations ^c	14.52	18.67	9.41	6.91	1.32	48
Days at the RMRU during the study	64.52	60.63	247.29	139.67	8.20**	116

^a Unit excluding sample $n = 50$; ^b subsample $n = 34$; ^c Missing : (Unit : 27, subsample : 7)

** $p < 0.001$

aggression, verbal aggression, aggression against objects, sexual aggression, self-neglect, substance use, unauthorized leave, stalking, suicidality, self-harm and victimization. The SOS behaviors are rated from 1 (minor) to 4 (severe), however, for the purpose of the current study, only the dichotomous (yes/no) presence of the behavior was analyzed. The interrater reliability of the SOS when coded from files has been found to be adequate ($ICC_2 = 0.70$).

START. The START is a structured clinical guide for the assessment and management of risk. Evaluators rate 20 dynamic items ranging from social adjustment, physical and mental health to treatment adherence in order to set up an intervention plan. The 20 items are scored both as vulnerabilities and as strengths, each on a three-point scale. The 20 items are then used to guide clinical staff to rate the seven risk estimates as either low, medium, or high risk (risk to others, self-harm, suicidality, unauthorized leave, self-neglect, substance abuse, and victimization by others). Initial studies on the START's psychometric properties indicate good interrater reliability (Cronbach's alpha .87), and predictive validity (AUCs of .65 to .77) for challenging behaviors (Nicholls et al., 2006). The START has recently been translated into French (Crocker et al., 2007) and both the English and French versions of the START are used by staff at the Douglas Institute.

Procedure

Permission was obtained from the Director of professional services of the hospital to consult patient files. The research protocol and procedures were approved by the Douglas Hospital Research Ethics Board. All data were denormalized to conduct analyses. Staff of the unit were trained to use the START by the head psychologist of the unit (D.G.), a former Ph.D. student (T.J.) and the nurse program coordinator (C.V.) who all had previously received training by one of the developers of the START and A.G.C. Staff used both the original English (Webster et al., 2004) and French versions of the START (Crocker et al., 2007) depending on what

language they felt most comfortable with. Trained research assistants who were blind to the START assessments coded the SOS according to the patient files.

Statistical Analyses

For the purpose of the analyses, four challenging behaviors (physical aggression, verbal aggression, sexual aggression, and stalking) were aggregated to create an "aggression against others" outcome variable. The other challenging behaviors that have a corresponding risk estimates on the START were analyzed separately and independently (i.e., self-harm, suicidality, unauthorized leave, substance use, self-neglect and victimization). Only aggression against objects of the SOS does not map directly on to the seven risk estimates of the START. Therefore, this outcome was omitted from the present analyses.

Phi correlations were calculated to examine the association between the occurrence of a challenging behavior prior to and after the START assessment. Spearman correlations were calculated to examine the association between the strength and vulnerability scores on each item.

Receiver Operating Characteristic (ROC) analyses were used to examine the START's predictive accuracy for the seven behavior categories in the 30 days post evaluation. ROC analysis has become increasingly common in violence prediction research due to its independence of base rates (Rice & Harris, 1995). The area under the curve statistic (AUC) represents the probability of randomly selecting an individual who has a challenging behavior and scores higher on the vulnerabilities and inverted strength scales, compared to an individual without a challenging behavior (Swets, 1992). An AUC of .50 indicates accuracy is equal to chance while an AUC of 1.00 indicates perfect accuracy. ROC analyses were conducted with ROCTools (Allaire & Cismaru, 2007). The Delong, Delong, and Clark-Pearson (1988) test was used to allow the statistical comparison of two ROC curves originating from the

same sample. The test provides a Chi-squared statistic and its associated significance at .05.

It was clear that not all items of the START were related to all of the challenging behaviors. Preliminary analyses (available upon request) allowed us to further construct “optimized scales” using items that were significantly associated with the occurrence of a challenging behavior occurring 30 days after the START assessment. See Table 2 for a list of the items of the optimized scales. The goal was to compare the original scale scores with the optimized scale scores and see if there was a significant improvement in prediction for each of the seven challenging behaviors.

Multilevel logistic regressions, using HLM (Raudenbush & Bryk, 2001), were estimated using START strength and vulnerability scales to predict a

Table 2 : Items included in the optimized scales

Optimized Scale	Items	
	Vulnerabilities	Strengths
Aggression towards others	Mental State	Recreational
	Impulse control	Mental State
	External Triggers	Impulse Control
	Conduct	External Triggers
		Rule Adherence
		Conduct
Self-Harm	Mental State	Mental State
		Rule Adherence
Suicidality	Recreational	Recreational
	Emotional State	Medication Adherence
	External Trigger	Coping
	Conduct	
	Coping	
Unauthorized Leave	Self-Care	Substance Use
	Substance Use	External Triggers
	Medication Adherence	
	Rule Adherence	
	Treatability	
Substance Abuse	Substance Use	Substance Use
	Insight	Treatability
Self-Neglect	Self Care	Social Skill
		Self Care
		External Trigger
		Conduct
Victimization	Emotional State	Impulse Control
	Impulse Control	Rule Adherence
	External Trigger	Conduct
	Attitudes	
	Rule Adherence	
	Conduct	

dichotomous event (a challenging behavior) in the 30 days after the evaluation. The advantage of using this technique is that it allowed for varying numbers of START evaluations per person, and the occurrence of an event 30 days prior the START was statistically controlled for without violating the assumption of independent observations. This allowed for the effect of the strength and vulnerabilities scales to be examined, above and beyond the influence of recent behaviors. Five sets of multilevel logistic regressions were conducted with the following predictors for each of the seven dichotomous behavior categories: 1) risk estimates, 2) vulnerability total scores, 3) strength total scores, 4) optimized vulnerability scale, and 5) optimized strength scale.

RESULTS

Among all patients admitted on the unit, a total of 74 (87%) patients had at least one challenging behavior. The average number of challenging behaviors per person was 23.85 (during the two year period of the study). For the subsample of patients ($n = 34$), base rates of engaging in any challenging behavior were 85% 30 days before the START evaluation, 91% 30 days after the START evaluation and 82% occurring in the 30 days prior to and after the START. Table 3 contains frequencies for each category of challenging behavior before and after the START assessment for the subsample of patients. This table illustrates the number of STARTs that were used in the analyses for each category of behavior. For instance, there were 70 out of 133 (53%) START assessments that had a physical aggression occurring 30 days before the evaluation. This corresponds to 25 patients out of 34 (74%) that displayed physical aggression occurring 30 days prior to the START. The total number of START assessments per behavior does not cumulate to 133 in cases where the behavior occurred on the same day as the assessment, and therefore it was impossible to categorize it as being before or after the START. Phi correlations (see Table 3) confirm that the presence of a behavior before the START

Table 3 : Base rates of challenging behaviors per individual and per START

Type of behavior	30 days before				30 days after				START Phi
	STARTa		Patientb		STARTa		Patientb		
	n	%	n	%	n	%	n	%	
Aggression toward others	70	52.6	25	73.5	71	53.4	24	70.6	.44**
Self-harm	8	6.0	7	20.6	10	7.5	8	23.5	.17
Suicide	14	10.5	7	20.6	14	10.5	7	20.6	.28*
Unauthorized leave	28	21.1	15	44.1	28	21.1	13	38.2	.23*
Substance use	30	22.6	11	32.4	31	23.3	13	38.2	.51**
Self-Neglect	42	31.6	17	50.0	47	35.3	20	58.8	.45**
Victimization	24	18.0	13	38.2	32	24.1	18	52.9	.24*

a n = 133; b n = 34; * p < 0.05, ** p < 0.001

assessment was significantly correlated with the presence of a behavior after the START in the case of aggression toward others, suicidality, unauthorized leave, substance use, self-neglect and victimization, but not self-harm. For this reason, past behavior was controlled for in subsequent prediction analyses in multilevel logistic regressions.

Table 4 presents the means for all the START items, on both the strength and vulnerability scales. The strength scale had a mean of 14.84 (SD=6.43) while the vulnerability scale had a mean of 21.23 (SD = 6.85). Each of the items' vulnerabilities and strengths correlate significantly (correlations ranging from -.36 to -.90), which resulted in collinearity between the two scales, $r = .89$. For this reason, both scales were tested separately in all multilevel logistic regression models.

Table 4 also illustrates that the correlation coefficients within the items were not homogeneous. For instance, material resources, whether coded as a strength or a vulnerability seemed to be tapping into the same construct. Whereas, the item relationships did not appear to be coded as the mirror inverse when rated as a strength or a vulnerability.

Results from the ROC analyses are presented in Table 5. The strength scales were inverted to facilitate interpretation. The vulnerability scale significantly predicted the occurrence of aggression towards others (AUC=.66, $p < .05$), unauthorized leave (AUC=.65, $p <.05$) and substance use (AUC = 67, $p <.05$) within 30 days after the START assessment. The strength scale also significantly

predicted the occurrence of aggression towards others (AUC = .65, $p < .05$), unauthorized leave (AUC=.65, $p <.05$) and substance use (AUC= .63, $p < .05$). Neither strength nor vulnerability total scores significantly predicted the occurrence of self-harm, suicidality, self-neglect or victimization (AUCs range from .52–.58, $p > .05$). For the risk estimates, only the item substance abuse significantly predicted the occurrence of substance use (AUC = .78, $p <.05$).

The optimized vulnerability scales (see Table 2) significantly predicted aggression towards others (AUC = .70, $p < .05$), suicidality (AUC = .67, $p < .05$), unauthorized leave (AUC = .67, $p < .05$), substance use (AUC = .78, $p < .05$) and self-neglect (AUC = .64,

Table 4 : Means and standard deviations and intercorrelations of vulnerabilities and strengths per START item

START Items	Strength	Vulnerabilities	Spearman's
	M (SD)	M (SD)	Rho
Social skills	0.75 (0.69)	1.06 (0.59)	-.66**
Relationships	0.51 (0.59)	1.05 (0.71)	-.36**
Occupational	0.35 (0.54)	1.55 (0.73)	-.87**
Recreational	0.66 (0.76)	1.17 (0.76)	-.80**
Self-care	1.25 (0.68)	0.68 (0.66)	-.77**
Mental state	0.44 (0.56)	1.39 (0.68)	-.80**
Emotional	0.98 (0.58)	0.74 (0.64)	-.53**
Substance use	1.35 (0.81)	0.66 (0.85)	-.87**
Impulse control	1.02 (0.72)	0.65 (0.70)	-.66**
External triggers	0.95 (0.68)	0.77 (0.76)	-.70**
Social support	0.45 (0.63)	1.55 (0.62)	-.67**
Material resources	0.19 (0.58)	1.81 (0.57)	-.90**
Attitudes	0.77 (0.56)	0.67 (0.67)	-.56**
Medication	1.16 (0.65)	0.88 (0.81)	-.67**
Rule adherence	1.14 (0.68)	0.68 (0.67)	-.71**
Conduct	0.88 (0.65)	0.51 (0.68)	-.67**
Insight	0.34 (0.54)	1.58 (0.57)	-.65**
Plans	0.41 (0.58)	1.45 (0.65)	-.73**
Coping	0.67 (0.57)	1.02 (0.65)	-.54**
Treatability	0.59 (0.59)	1.38 (0.66)	-.74**
Total Score	14.84 (6.43)	21.23 (6.85)	-.89**

n =133 START evaluations; * p < 0.05, ** p < 0.001

Table 5 : ROC AUC's of the original and optimized strength and vulnerability scales, risk estimates and the occurrence of an event 30 days after the START assessment

	Types of behaviors						
	Aggression toward others	Self-Harm	Suicidality	Unauthorized Leave	Substance Use	Self-Neglect	Victimization
	AUC (95% CI)	AUC (95% CI)	AUC (95% CI)	AUC (95% CI)	AUC (95% CI)	AUC (95% CI)	AUC (95% CI)
Original START scales							
Vulnerabilities	0.66 (.56-.75)**	0.58 (.20-.75)	0.58 (.46-.70)	0.65 (.53-.78)*	0.67 (.55-.78)*	0.52 (.41-.62)	0.55 (.43-.66)
Strengths ^a	0.65 (.56-.74)*	0.57 (.38-.76)	0.57 (.45-.70)	0.66 (.53-.78)*	0.63 (.51-.75)*	0.53 (.42-.63)	0.53 (.42-.65)
Risk Estimates	0.52 (.42-.62)	0.54 (.34-.73)	0.52 (.36-.67)	0.55 (.42-.68)	0.78 (.70-.87)**	0.55 (.45-.65)	0.51 (.39-.62)
Optimized Scales							
Vulnerabilities	0.70 (.61-.79)**	0.62 (.46-.78)	0.67 (.57-.76)*	0.67 (.56-.79)*	0.78 (.68-.88)**	0.64 (.54-.74)*	0.62 (.52-.74)*
Strengths ^a	0.65 (.56-.75)**	0.71 (.56-.85)*	0.78 (.66-.90)**	0.70 (.58-.82)**	0.78 (.68-.89)**	0.66 (.56-.76)*	0.59 (.47-.71)
Difference between the optimized and original scales b							
Vulnerabilities	1.50(1)	0.21(1)	4.30(1)*	0.18(1)	5.77(1)*	5.06(1)*	15.91(1)**
Strengths	0.02(1)	3.10(1)	7.84(1)*	0.84(1)	11.80(1)**	22.02(1)**	8.87(1)*

n = 133; a Concave curves have been inverted; b using Delong, Delong & Clarke-Pearson (1988) method; $\chi^2(df)$

* p < 0.05, ** p < 0.001

p < .05). The optimized strength scales significantly predicted violence to others (AUC = .65, p < .05), self-harm (AUC = .71, p < .05), suicidality (AUC = .78, p < .05), unauthorized leave (AUC = .70, p < .05), substance use (AUC = .78, p < .05) and self-neglect (AUC = .66, p < .05). A comparison of two AUCs using the Delong et al. (1988) method (see Table 5) showed that the optimized vulnerability scale was significantly better at predicting suicidality, substance abuse, self-neglect, and victimization than the original vulnerability scale. The optimized strength scale was significantly better at predicting suicidality, substance, self-neglect, and victimization than the original strength scale.

Table 6 contains the results from the multilevel logistic regressions for the original vulnerability, strength and the optimized scales for each of the seven challenging behavior outcomes. Neither the original strength nor the vulnerability total scores predicted the presence of a challenging behavior, in any category, above and beyond the influence of a past behavior, with odds ratios (ORs) ranging from 0.96 to 1.05. The optimized vulnerability scale significantly predicted aggression towards others, self-harm, substance abuse and victimization. For every one unit increase in the optimized vulnerability scale, there was an increase in odds of aggression towards others (OR = 1.23), self-harm (OR = 2.78), substance abuse (OR = 1.89) and victimization (OR = 1.26).

Table 6 : Hierarchical logistic regressions for strengths, vulnerabilities and the optimized scales

	Types of behaviors						
	Aggression towards others	Self-Harm	Suicide	Unauthorized Leave	Substance Abuse	Self-Neglect	Victimization
	O.R. (95% CI)	O.R. (95% CI)	O.R. (95% CI)	O.R. (95% CI)	O.R. (95% CI)	O.R. (95% CI)	O.R. (95% CI)
Original START scales							
Intercept	0.53 (0.31-0.91)*	0.07 (0.03-0.14)**	0.11 (0.05-0.23)**	0.21 (0.13-0.34)**	0.15 (0.08-0.28)**	0.28 (0.17-0.46)**	0.26 (0.14-0.48)**
Behavior before	4.70 (2.66-8.32)**	3.59 (0.77-16.62)	2.05 (0.40-10.62)	1.48 (0.37-5.98)	7.96 (2.37-26.77)*	6.29 (2.99-13.20)**	2.70 (0.94-7.79)
Vulnerabilities	1.05 (0.99-1.11)	1.05 (0.97-1.13)	1.00 (0.95-1.05)	1.04 (0.96-1.14)	0.99 (0.94-1.06)	1.01 (0.96-1.07)	1.05 (0.99-1.12)
Intercept	0.52 (0.30-0.92)*	0.07 (0.03-0.14)**	0.10 (0.05-0.22)**	0.20 (0.12-0.33)**	0.14 (0.08-0.27)**	0.28 (0.17-0.46)**	0.26 (0.15-0.48)**
Behavior before	4.85 (2.67-8.81)**	4.10 (0.82-20.62)	2.09 (0.41-10.70)	1.67 (0.40-7.03)	8.83 (2.50-31.18)*	6.34 (2.97-13.55)**	2.63 (0.95-7.31)
Strengths	0.96 (0.90-1.02)	0.96 (0.88-1.05)	0.99 (0.92-1.05)	0.97 (0.88-1.06)	1.03 (0.95-1.11)	0.99 (0.93-1.06)	0.97 (0.91-1.03)
Optimized Scales							
Intercept	0.55 (0.31-0.99)*	0.06 (0.03-0.13)**	0.10 (0.05-0.22)**	0.21 (0.13-0.35)**	0.15 (0.08-0.29)**	0.28 (0.17-0.45)**	0.23 (0.12-0.45)**
Behavior before	4.40 (2.34-8.27)**	3.06 (0.52-17.93)	2.12 (0.41-10.82)	1.22 (0.33-4.41)	4.84 (1.26-18.56)*	6.24 (3.00-12.96)**	2.75 (0.92-8.23)
Vulnerabilities	1.23 (1.03-1.45)*	2.78 (1.19-6.50)*	1.03 (0.89-1.19)	1.24 (0.96-1.59)	1.89 (1.30-2.76)*	0.51 (0.25-1.04)	1.26 (1.11-1.44)*
Intercept	0.54 (0.30-0.98)*	0.05 (0.03-0.10)**	0.07 (0.03-0.15)**	0.21 (0.12-0.35)**	0.16 (0.08-0.30)**	0.28 (0.17-0.49)**	0.24 (0.13-0.47)**
Behavior before	4.53 (2.41-8.53)**	3.75 (0.63-22.14)	2.10 (0.47-9.36)	1.16 (0.36-3.74)	4.70 (1.05-20.88)*	5.71 (2.86-11.37)**	2.42 (0.83-7.08)
Strengths	0.89 (0.78-1.02)	0.38 (0.21-0.68)*	0.47 (0.29-0.74)*	0.60 (0.41-0.88)*	0.63 (0.38-1.03)	0.75 (0.56-1.02)	0.72 (0.56-0.92)*

n = 133 START evaluations nested in n = 34 patients; OR = Odds Ratio; * p < 0.05, ** p < 0.001

The optimized vulnerability scale did not significantly predict suicidality, unauthorized leave, or self-neglect. The optimized strength scale significantly predicted self-harm, suicidality, unauthorized leave, and victimization. For every unit increase in the optimized strength scale, there was a decrease in the odds of self harm (OR = 0.38), suicidality (OR = 0.47), unauthorized leave (OR = 0.60) and victimization (OR = 0.72). The optimized strength scale did not significantly predict aggression towards others, substance use or self-neglect.

DISCUSSION

The current study examined the START's predictive validity for multiple types of risks. The results of the present study lend partial support to the use of the START in the assessment of risk of certain types of challenging behaviors and shed light into the manner in which the START is being used in daily clinical practice.

One of the benefits of the START is that each item can be coded as both a strength and a vulnerability. However, to date, no studies have investigated the utility of such a rating system. Upon examination of the correlations between the strength and vulnerability scores of each item, items such as relationships ($r = -.36$), emotional state ($r = .53$), and coping ($r = -.54$) appear to be readily considered as strengths and vulnerabilities. For instance, an individual can have successful coping strategies in one domain, and yet lack in others. Conversely, material resources ($r = -.90$) are not easily conceptualized on both dimensions and therefore having this item rated on both scales may not be useful.

The ROC results showed that, when recent behavior was not controlled for, both the strength and vulnerability scales significantly predicted aggression against others, suicidality and substance abuse. However, the risk estimates, rated as high, medium or low for each of the seven challenging

behaviors, did not perform as well. The only risk estimate that was significantly predictive of challenging behavior was substance use in predicting the occurrence of substance use. It may be that the clinical team was more comfortable with the items when completing the ratings, than the risk estimates. However, in a study that used similar outcomes of challenging behaviors (McNeil & Binder, 1991), the authors found that physicians and nurses using probabilistic estimates (high, medium, and low) were able to accurately predict the occurrence of an event one week following admission. The McNeil and Binder study lends support to the use of categorical probabilistic risk estimates, despite the fact that nurses and physicians were not using any kind of structured guide to aid their decisions.

The large quantity of items on the START may render the evaluation and prediction of specific behaviors difficult, and not all items are intuitively related to each type of challenging behavior. This finding was confirmed by the multilevel logistic regressions, where only specific items were related to specific behaviors. The optimized scales allowed us to predict each outcome using only a limited number of items that correlated with the outcome. These optimized vulnerability and strength scales were able to accurately predict a greater number of types of challenging behaviors than the original scales. Still, the optimized scales' AUC's were only significantly better (for both vulnerabilities and strengths) in the case of suicidality, substance use, self-neglect, and victimization. It is clear that these optimized scales cannot be generalized beyond the present sample and must be tested on a validation sample to eliminate the influence of sample specific variability.

Previous research with the START has tended to use total vulnerability and risk scores to examine predictive validity. However, issues relating to the relevance of specific items in risk assessment schemes have begun to attract attention from scholars (Douglas, 2009). Specifically, rather than focusing on the simple presence or absence of a risk

factor, a shift toward the relevance of items (and their combination) lends well to the idea of dynamic and multiple risks. Upon examination of the items comprising the optimized scales, it is clear that some items intuitively and statistically relate to particular outcomes (e.g., impulse control and aggression towards others). More work is needed on how combinations of START items predict various types of adverse outcomes.

In the current study, once recent past behavior was controlled for, the strength and vulnerabilities scales no longer predicted challenging behaviors. This may indicate that mental health professionals who complete the START are reflecting on past challenging behaviors to complete the START ratings more than reflecting on the items. This finding is not entirely surprising given that the START is a clinical instrument and clinicians evaluate current situations based on past events and observations. Furthermore, it would be surprising that clinicians not use recent past behavior to influence their scores on specific items. Conversely, it may be that successful risk management interventions were implemented which would explain why past behaviors frequently correlated with START scores, but not future behaviors.

It should be noted that the fact that not all items predicted all challenging behaviors does not detract from one of the START's main benefit: to provide guidance in clinical practice (Grevatt et al., 2004). The interaction between the clinical teams who conduct the assessments, and the fact that they are responsible for preventing the very behaviors they are predicting is an inevitable limitation in this type of research (McNeil & Binder, 1991).

Limitations

Unlike many studies that use risk assessment schemes, these results can be interpreted as an indication of how clinical teams actually use the START in routine practice. Another strength of this research was that past behavior was controlled for in the multilevel logistic models. Despite the adage that

past behavior is the best predictor of future behavior and that this has been replicated in research done with the severely mentally ill (Amore et al., 2008), many studies do not take it into account in a systematic manner (e.g., HCR-20, Webster et al., 1997).

This was the first attempt to examine the predictive validity of clinical ratings of the START. The present study was hindered by the fact that the strength and vulnerabilities scores were highly collinear ($r = -.89$), and therefore could not be tested concomitantly. As a result, no mediation or moderator effects could be tested. Further research could test a mediational model where protective factors could presumably change the strength of the relationship between risk factors and challenging behaviors. In fact, Rogers (2000) suggested that inpatient risk assessments are likely to be influenced by mediation effects. Douglas and Skeem (2005) also noted the complexity of this issue in saying, "it is often difficult to discern the relation between these risk factors and outcomes. The relation could be direct, moderated or mediated by a third variable, or altogether caused by a third variable" (p. 351).

The present results should also be interpreted carefully given that some confidence intervals of the AUC's and odds ratios were large. This is a result of a small sample size producing large standard errors. It is clear that this study needs to be replicated on a larger sample.

Another limitation of the current study was that interventions occurring subsequent to incidents were not controlled for. Presumably, successful interventions decrease the likelihood and/or severity of future challenging behaviors. Future research should examine how risk management interventions following problematic behavior affect their occurrence.

More work should focus on examining how behavior severity correlates with the items and risk estimates in order to clarify the relation between the START and multiple risks. Further analysis of the

severities of aggression would have been informative, but impossible with the current dataset given the low base rate of severe behaviors and the high base rates of mildly challenging behaviors.

Conclusion

The present research represents an important step in the implementation of risk assessment schemes in clinical practice. The shift towards structured professional guides for the assessment and management of risk is crucial, but the work does not stop there. The way in which clinical teams use these tools must be studied, and then further improved to tailor to their needs. Risk assessment instruments need to be validated in the environments where they are intended to be used, and with the professionals that they are intended to be used by.

REFERENCES

- Allaire, J.-F., & Cismaru, D. (2007). ROCTools: A statistical software for the analysis of ROC curves (Version 1.0.2). Montreal, Canada: The Statistical Consulting Group of Centre de recherche de l'Institut Philippe-Pinel de Montréal.
- Almvik, R., Woods, P., & Rasmussen, K. (2000). The Broset Violence Checklist: Sensitivity, specificity, and interrater reliability. *Journal of Interpersonal Violence*, 15, 1284–1296.
- Amore, M., Menchetti, M., Tonti, C., Scarlatti, F., Lundgren, E., Esposito, W., et al. (2008). Predictors of violent behavior among acute psychiatric patients: Clinical study. *Psychiatry and Clinical Neurosciences*, 62, 247–255.
- Barlow, K., Grenyer, B., & Ilkiw-Lavalle, O. (2000). Prevalence and precipitants of aggression in psychiatric inpatient units. *Australian and New Zealand Journal of Psychiatry*, 34, 967–974.
- Borum, R., Bartel, P., & Forth, A. (2006). Structured Assessment of Violence Risk in Youth (SAVRY). Odessa, FL: Psychological Assessment Resources.
- Brook, J. S., Whiteman, M., Gordon, A. S., & Cohen, P. (1989). Changes in drug involvement: A longitudinal study of childhood and adolescent determinants. *Psychological Reports*, 65 (3, Pt 1), 707–726.
- Choe, J. Y., Teplin, L. A., & Abram, K.M. (2008). Perpetration of violence, violent victimization, and severe mental illness: Balancing public health concerns. *Psychiatric Services*, 59, 153–164.
- Crocker, A. G., Braithwaite, E., Laferrière, D., Jenkins, T., Gagnon, D., & Venegas, C. (in press). START changing cultures: Implementing a risk assessment and management tool in a civil psychiatric setting. *International Journal of Forensic Mental Health*.
- Crocker, A. G., & Côté, G. (2009). Evolving systems of care: Individuals found not criminally responsible on account of mental disorder in custody in civil and forensic psychiatric settings. *European Psychiatry*, 24, 356–364.
- Crocker, A. G., Garcia, A., Israël, M., Hindle, Y., Gagnon, D., Venegas, C., et al. (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.
- Crocker, A. G., Grenier, M., Webster, C. D., Nicholls, T. L., Martin, M.-L., Desmarais, S., & Brink, J. (2007). 'Évaluation du risque à court terme et traitabilité: Un guide d'évaluation et de planification. Version française du manuel START (Webster et al.). Institut universitaire en santé mentale Douglas.
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44, 837–845.
- Desmarais, S. L., Nicholls, T. L., Read, J. D., & Brink, J. (2010). Confidence and accuracy in assessments of short-term risks presented by forensic psychiatric patients. *The Journal of Forensic Psychiatry & Psychology*, 21, 1–22.
- de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbé, M. (2009). SAPROF: Guidelines for the assessment of protective factors for violence. Utrecht, The Netherlands: Forum Educatief.
- Douglas, K. S. (2009). What should the next generation of violence risk assessment look like? Building on current knowledge to advance the field. Keynote presentation at the 9th annual conference of the International Association of Forensic Mental Health Services, Edinburgh, Scotland.
- Douglas, K. S., & Ogloff, J. R. (2003). Multiple facets of risk of violence: The impact of judgmental specificity on

- structured decisions about violence risk. *International Journal of Forensic Mental Health*, 2, 19–34.
- Douglas, K. S., Ogloff, J. R., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology*, 67, 917–930.
- Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law*, 11, 347–383.
- Doyle, M., & Dolan, M. (2002). Violence risk assessment: Combining actuarial and clinical information to structure clinical judgment for the formulation and management of risk. *Journal of Psychiatric and Mental Health Nursing*, 9, 649–657.
- Doyle, M., & Dolan, M. (2008). Understanding and managing risk. In K. Soothill, P. Rogers & M. Dolan (Eds.), *Handbook for forensic mental health* (pp. 244–266). Cullompton: William Publishing.
- Doyle, M., Lewis, G., & Brisbane, M. (2008). Implementing the Short-Term Assessment of Risk and Treatability in a forensic mental health service. *Psychiatric Bulletin*, 32, 406–408.
- Duxbury, J. (2002). An evaluation of staff and patient views of and strategies employed to manage inpatient aggression and violence on one mental health unit: A pluralistic design. *Journal of Psychiatric & Mental Health Nursing*, 9, 325–337.
- Ferris, L. E., Sandcock, J., Hoffman, B. F., Silverman, M. M., Barkun, H., Carlisle, J., et al. (1997). Risk assessment for acute violence to third parties: A review of the literature. *Canadian Journal of Psychiatry*, 42, 1051–1060.
- Gagliardi, G. J., Lovell, D., Peterson, P. D., & Jemelka, R. (2004). Forecasting recidivism in mentally ill offenders released from prison. *Law and Human Behavior*, 28, 133–155.
- Grevatt, M., Thomas-Peter, B., & Hughes, G. (2004). Violence, mental disorder and risk assessment: Can structured clinical assessments predict the short-term risk of inpatient violence? *Journal of Forensic Psychiatry & Psychology*, 15, 278–292.
- Hodgins, S., Alderton, J., Cree, A., Aboud, A., & Mak, T. (2007). Aggressive behaviour, victimisation and crime among severely mentally ill patients requiring hospitalisation. *British Journal of Psychiatry*, 191, 343–350.
- Joyal, C. C., Gendron, C., & Côté, G. (2008). Nature and frequency of aggressive behaviours among long-term inpatients with schizophrenia: A 6-month report using the Modified Overt Aggression Scale. *La Revue canadienne de psychiatrie*, 53, 478–481.
- Lodewijks, H. P. B., Doreleijers, T. A. H., De Ruiter, C., & Borum, R. (2008). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during residential treatment. *International Journal of Law and Psychiatry*, 31, 263–271.
- McNeil, D. E., & Binder, R. L. (1991). Clinical assessment of the risk of violence among psychiatric inpatients. *American Journal of Psychiatry*, 148, 1317–1321.
- McNeil, D. E., & Binder, R. L. (1995). Correlates of accuracy in the assessment of psychiatric inpatients' risk of violence. *American Journal of Psychiatry*, 152, 901–906.
- McNeil, D. E., Gregory, A. L., Lam, J. N., Binder, R. L., & Sullivan, G. R. (2003). Utility of decision support tools for assessing acute risk violence. *Journal of Consulting and Clinical Psychology*, 71, 945–953.
- Menckel, E., & Viitasara, E. (2002). Threats and violence in Swedish care and welfare: Magnitude of the problem and impact on municipal personnel. *Scandinavian Journal of Caring Sciences*, 16, 376–385.
- Miller, H. A. (2006). A dynamic assessment of offender risk, needs and strengths in sample of pre-release general offenders. *Behavioral Sciences and the Law*, 24, 767–782.
- Narrow, W., Regier, D., Rae, D., Manderscheid, R., & Locke, B. (1993). Use of services by persons with mental and addictive disorders: Findings from the National Institute of Mental Health Epidemiologic Catchment Area Program. *Archives of General Psychiatry*, 50, 95–107.
- Newcomb, M. D., & Felix-Ortiz, M. (1992). Multiple protective and risk factors for drug use and abuse: Cross-sectional and prospective findings. *Journal of Personality and Social Psychology*, 63, 280–296.
- Nicholls, T. L., Brink, J. H., Desmarais, S. L., Webster, C. D., & Martin, M.-L. (2006). The Short-Term Assessment of Risk and Treatability (START) A prospective validation study in a forensic psychiatric sample. *Assessment*, 13, 313–327.
- Nicholls, T. L., Gagnon, N., Crocker, A. G., Brink, J. H., Desmarais, S. L., & Webster, C. D. (2007). START Outcome Scale SOS. Vancouver, Canada: BC Mental Health & Addiction Services.

- Nijman, H. (1999). Aggressive behavior of psychiatric inpatients: Measurement, prevalence, and determinants. Maastricht, the Netherlands: University Press Maastricht.
- Nijman, H., Bowers, L., Oud, N., & Jansen, G. (2005). Psychiatric nurses' experience with inpatient aggression. *Aggressive Behavior*, 31, 217–227.
- Ogloff, J. R., & 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.
- Raudenbush, S. W., & Bryk, A. S. (2001). Multilevel linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage.
- Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, 63, 737–748.
- Rogers, R. (2000). The uncritical acceptance of risk assessment in forensic practice. *Law and Human Behavior*, 24, 595–605.
- Rutter, M. (1987). Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry*, 5, 316–331.
- Sheldrick, C. (1999). Practitioner review: The assessment and management of risk in adolescents. *Journal of Child Psychology and Psychiatry*, 40, 507–518.
- Steadman, H. J., Monahan, J., Robbins, P. C., Appelbaum, P., Grisso, T., Klassen, D., et al. (1993). From dangerousness to risk assessment: Implications for appropriate research strategies. In S. Hodgins (Ed.), *Mental disorder and crime* (pp. 39–62). Thousand Oaks, CA: Sage.
- Stübner, S., Groß, G., & Nedopil, N. (2006). Inpatient risk management with mentally ill offenders: Results of a survey on clinical decision-making about easing restrictions. *Criminal Behaviour and Mental Health*, 16, 111–123.
- Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist*, 47, 522–532.
- Towl, G. (2005). Risk assessment. *Evidence-Based Mental Health*, 8, 91–93.
- Wang, E. W., & Diamond, P. M. (1999). Empirically identifying factors related to violence risk in corrections. *Behavioral Sciences & the Law*, 17, 377–389.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). HCR-20: Assessing risk of violence (Version 2). Burnaby, Canada: Mental Health Law and Policy Institute, Simon Fraser University.
- Webster, C. D., Martin, M.-L., Brink, J. H., Nicholls, T. L., & Middleton, C. (2004). Short-Term Assessment Risk and Treatability (START). St. Joseph's Healthcare, BC: Forensic Psychiatric Services Commission.
- Yudofsky, S. C., Silver, J. M., Jackson, W., Enticott, J., & Williams, D. (1986). The Overt Aggression Scale for the objective rating of verbal and physical aggression. *American Journal of Psychiatry*, 143, 35–39.