THE RELIABILITY AND VALIDITY OF THE SAPROF AMONG FORENSIC MENTAL HEALTH PATIENTS

by

Sandra Oziel
Master of Arts in Psychology, Ryerson University 2012

A dissertation

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Program of

Psychology

Toronto, Ontario, Canada, 2016

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A DISSERTATION

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this dissertation to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my dissertation may be made electronically available to the public.

The Reliability and Validity of the SAPROF among Forensic Mental Health Patients

Sandra Oziel

Doctor of Philosophy, 2016
Psychology
Ryerson University

Abstract

Assessing and managing level of risk among forensic mental health patients is a primary role of clinical forensic psychologists. Forensic assessments are focused on risk factors and deficits, whereas patient strengths and protective factors are either partially included or overlooked altogether by forensic psychologists. As a result, less is known about protective factors in general and how they may serve to inform risk management practices. The Structured Assessment of Protective Factors for Violence Risk (SAPROF) is the first tool to exclusively rely on protective factors and was investigated for the current study.

The psychometric properties of the SAPROF were examined using a sample of 50 Canadian patients found Not Criminally Responsible (NCR) at a psychiatric hospital using both file information and semi-structured interviews. Outcome variables included risk management decisions (change in privilege level and security level) and indicators of recidivism (psychiatric medication administration, institutional misconduct and disposition breaches). The study found some evidence for intrarater and interrater reliability, construct validity, predictive validity and incremental predictive validity. The SAPROF approached significance for adding incremental predictive validity to the HCR-20 V3, a measure of violence risk, for disposition breaches and institutional misconduct, and effect sizes doubled.

Given that the addition of the SAPROF increased the accuracy of the violence risk assessment, there are considerable implications for informing clinical practice. Implications for risk assessment, treatment planning, intervention and risk management decisions implemented by review boards and clinical practitioners are discussed. It is recommended that the SAPROF be added as an adjunct measure to risk assessment batteries and included in hospital reports, given that it predicted several patient behaviours.

Acknowledgments

There are several people that made this study possible. I would like to thank my graduate supervisor Dr. David Day for his ongoing support and guidance. You are a true mentor that cared about both my academic success and personal growth. Thank you Dr. Lisa Marshall, you were instrumental in helping to develop and execute this study and it could not have been done without your expertise in risk assessment. Thank you Dr. Todd Girard, your psychometric and statistical guidance were essential to this study. Additionally, I would like to thank my examination committee members Drs. Alasdair Goodwill and Beth McCay, as well as Dr. Sandy Simpson for serving as my external examiner.

I would like to acknowledge the research department at Ontario Shores, specifically Michael Wasdell, Steve Mann, Tina Powers and Sampath Sridhara. I am also appreciative of the recruitment assistance provided by the talented team of clinical forensic psychologists at Ontario Shores. Drs. Erica Martin, Justine Joseph, Joanne Coutts Laura Leong, Marc Levi, Lara Hiseler, Krystle Martin, and Shelly McAlister, thank you for the time and effort you invested in this study despite your very busy schedules. Also, thank you to Sarah Cheney for your assistance in recruiting outpatients at the hospital.

In addition, thank you Dr. Michiel de Vries Robbe for your training, support and prompt replies to my complicated methodological questions. I am inspired by your work and hope others will continue researching this important topic. I would also like to thank John Zawadzki at CAMH for providing assessment measure materials. Thank you to Natalie Whitney for your assistance with data input.

I am grateful for the support and input provided by my peers, lab mates and friends, Saeid Chavoshi, Skye Stephens and Jennifer Newman. Thank you to my parents and grandmother for their encouragement; you will be happy to know that I am no longer a "professional student." Lastly, I am thankful to Aeran Kurzfeld for his unconditional support, endless hours of editing and patiently listening to me vent. You motivated and believed in me, we did it!

Table of Contents

List of Tables

List of Figures

Introduction

The inclusion of protective factors in forensic risk assessment has been described as a 'new frontier' in forensic mental health (de Ruiter & Nicholls, 2011). Protective factors are the characteristics of a person or environment that reduce the risk of future violent behaviours such as attitude towards authority and motivation to attend treatment (de Vogel, de Ruiter, Bouman, & de Vries Robbe, 2012). Previously, assessments have focused on risk factors and deficits, whereas patient strengths and protective factors have either been included partially or overlooked altogether by forensic psychologists (Rogers, 2000). As a result, less is known about protective factors and how they may serve to inform risk management practices (de Vogel, de Vris Robbe, de Ruiter, & Bouman, 2011). The identification and inclusion of protective factors in clinical services align with the Good Lives Model (GLM) of offender rehabilitation (GLM; Ward, 2002 a,b), a prominent framework that emphasizes strengths and capabilities to reduce reoffending (Ward, Mann, & Gannon, 2007).

The Structured Assessment of Protective Factors for Violence Risk (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbe, 2009) is the first tool to solely include protective factors to assess adults in conflict with the law, and can be used as an adjunct to other risk measures. There have been several empirical studies examining the reliability and validity of the SAPROF (de Vries Robbe, de Vogel, & de Spa, 2011; de Vries Robbe, de Vogel, & Douglas, 2013; de Vries Robbe, de Vogel, Douglas, & Nijman, 2015; de Vries Robbe, de Vogel, Koster, & Bogaerts, 2015). The SAPROF is a promising tool designed to inform risk prediction, risk management and treatment planning (de Vogel et al., 2011). Nevertheless, prior studies have been limited with

respect to the methodology and exclusion criteria implemented, impacting the generalizability of these studies. These issues are further elaborated on in the literature review.

The literature review is organized into the following six sections: 1) Good Lives Model (GLM); 2) Not Criminally Responsible Persons; 3) Generations of Risk Assessment Instruments; 4) Defining Protective Factors; 5) Instruments Examining Protective Factors; and 6) Study Rationale. The identification and inclusion of protective factors is aligned with the Good Lives Model of Offender Rehabilitation (GLM; Ward, 2002a,b), which emphasizes strengths and capabilities of the individual to reduce reoffending (Ward, Mann, & Gannon, 2007). This section is followed by a description of Not Criminally Responsible on account of Mental Disorder (NCRMD) persons, which is the population under investigation in the present study. The Ontario Review Board (ORB) process, the governing body responsible for managing public risk relating to NCRMD persons is also reviewed. Subsequently, a brief summary of the generations of risk assessment tools are provided, since an understanding of the history of risk assessment is necessary to gain an appreciation of the way protective factors fit within current risk evaluations. The types of factors commonly incorporated in assessment tools are defined: static risk, dynamic risk, and protective factors. Additionally, recent validation studies on the SAPROF and other protective factor measures are reviewed. Lastly, the rationale for the inclusion of protective factors within a forensic context is discussed, along with the limitations of the literature.

The purpose of the present study is to examine the psychometric properties of the SAPROF among forensic patients across several short-term outcomes. Various forms of reliability and validity are investigated. The methodology implemented is the first to use semi-structured interviews to score the SAPROF, adding to the ecological validity of the study. A primary purpose of the study is to determine whether the addition of the SAPROF increased the accuracy

of the violence risk assessment. This study has considerable clinical implications for risk assessment, treatment planning, intervention and risk management decisions implemented by review boards and clinical practitioners.

Good Lives Model

The Good Lives Model (GLM; Ward, 2002a,b) of offender rehabilitation is a strength-based, holistic framework that highlights the strengths and competencies of the individual, in conjunction with their risk factors, to forecast desistance from crime (Ward et al., 2007). The primary aim of the framework is to identify the function of offending behaviour and equip offenders with internal and external resources to live a meaningful and prosocial life (Ward, Yates, & Willis, 2012). According to this perspective, identifying and building upon protective factors is an effective strategy for offenders in order to lead a fulfilling life and reduce recidivism, rather than solely evaluating risk and imposing risk management strategies (Ward & Brown, 2004).

From the GLM perspective, inclusion of protective factors when predicting risk is essential for ensuring balanced decisions are made for managing patient risk, for both offenders and forensic mental health patients (Barnao, 2013; Ward, Mann, & Gannon, 2007). Assessment tools that provide a quantitative/actuarial risk estimation or assign individuals to risk categories, either partially address or completely disregard protective factors. Exclusively attending to negative attributes of patients and neglecting the individual's capabilities when assessing risk can have dire consequences, including assigning a higher level of risk than warranted, and increased restrictions of liberties.

The GLM is in line with the philosophy of positive psychology, most notably associated with the works of Seligman (2002). Seligman (2002) asserts that adopting a more constructive,

positive and strength-based approach to working with patients is essential within the field of psychology. The central reason for the positive psychology movement was to address the imbalance toward a negative perspective present in mainstream psychology (Wong, 2011). Specifically, Seligman emphasized that by attending to what is good about people may counteract the field of psychology's preoccupation with psychopathology (Seligman, 1998). As such, the "importance of amplifying clients' strengths rather than repair their weaknesses" should not be overlooked (Seligman, 2002, p. 5).

The GLM framework asserts that focusing on positive aspects of the individual results in increased patient engagement and motivation to change, because intervention goals are congruent with the personal goals and values of the individual. Therefore, although other models may strive to improve the offender's quality of life such as the Risk Need Responsivity framework (RNR) (Andrews, Bonta, & Hoge, 1990), it is not a primary objective, as it is with the GLM approach (Ward, 2002b).

According to the GLM, individuals are goal-directed and striving for the highest level of well-being by achieving "human/primary goods," which are basic human needs or life goals (Bouman, de Ruiter, & Schene, 2008; Ward, 2002b; Yates 2013). GLM describes primary goods as aspects of human functioning and experiences that benefit the individual and result in increased well-being (Ward, 2002b). As stated by Ward (2002b, p. 173), "the ultimate grounding of human striving or goals resides in primary goods: valued states of affairs, states of mind, characteristics, activities, or experiences that are sought for their own sake and that are derived from substantive facts about human nature." Examples of primary goods include: healthy living and optimal physical functioning, knowledge, excellence in play and work (e.g., mastery of experiences), excellence in agency (e.g., autonomy and self-directedness), inner peace (e.g.,

freedom from emotional turmoil and stress), friendship and community (e.g., relatedness to others, including intimate, romantic, and family relationships), spirituality (in the broad sense of finding meaning and purpose in life), happiness and creativity (Ward, 2002b; Ward & Stewart, 2003).

The way an individual prioritizes specific primary goods is a reflection of their personal values (Ward et al., 2012). Notably, each primary good is multi-faceted and there is no one preferred primary good valued by every person; rather it differs across individuals. Therefore, understanding the function of human behaviour, in terms of which primary goods are being acquired can be useful for comprehending motivations for behaviour.

Considering that most daily activities involve attaining basic primary goods, the same may also apply to the pursuit of antisocial goals (Ward, 2002b). According to the GLM framework, antisocial behaviour is indicative of the accumulation of actions exhibited in order to attain primary goods. Offending behaviour arises from achieving primary goods in socially unacceptable ways and possessing a mistaken belief about what is valuable and in the individual's best interest (Ward & Brown, 2004). As such, the main problem with antisocial behaviour lies in the "secondary means" or the approach used to secure goods (Barnao, 2013).

The concept of secondary means can be understood by examining how an individual strives to attain excellence at work (e.g., a primary good), for example. Instead of investing the time and effort to gain the necessary skills, they may plagiarize the works of others (e.g., secondary means). To further demonstrate this point and relate it to serious contact offences, an individual may desire intimacy with others and relatedness (e.g., primary good) but lack the skills to form a relationship and thus may sexually offend against women (e.g., secondary means) in order to attain feelings of closeness. Similarly, another individual may sexually offend

but does so to achieve a different primary good. For example, another offender may sexually offend against women in order to gain a sense of autonomy or power (e.g., primary good), which is obtained through sexual aggression (Lindsay, Ward, Morgan, & Wilson, 2007).

Motivation to commit antisocial behaviour may differ by individual and, without identifying the function of the behaviour, the problem cannot be effectively addressed. Also, these examples highlight that the problem may not lie in the valued primary goods, but rather the use of inappropriate strategies for achieving those goods (Ward & Brown, 2004). As such, an individual's criminogenic needs may be obstacles to achieving their primary goods (Bouman, de Ruiter, & Schene, 2008). Other examples of secondary means include suicide, stalking, arson, homicide, self-harm, violence, mental illness symptoms, and substance abuse (Barnao, Robertson, & Ward, 2010).

One critique of the GLM concerns the lack of explanation as to why offenders do not possess the primary goods they seek to begin with and why individuals engage in illegal behaviour (Serin & Lloyd, 2009). Additionally, little explanation is provided for the biological, ecological and neuropsychological variables responsible for offending behaviour (Serin & Lloyd, 2009).

GLM Assessment and Treatment. Evaluating an individuals' likelihood to reoffend or desist from crime using the GLM approach consists of assessing their goals. First, a comprehensive assessment is conducted to measure static and dynamic risk factors. Second, the individual identifies what they personally consider to be primary goods (Ward et al., 2012). This can be achieved by asking patients about their core commitments in life and daily activities and experiences (Ward et al., 2012). Also, enquiring about the patient's values that directly or indirectly contribute to their offending behaviour is useful (Ward et al., 2012).

Interventions incorporating both primary goods and reducing risk factors are considered most preferable, given the association between an individual's values, antisocial behaviour and rates of reoffending (Ward, 2002b; Ward & Brown, 2004). The GLM framework maintains that in order to effectively tackle criminogenic needs, it is necessary to understand an offender's goals or primary goods they are trying to achieve (Ward et al., 2012). Thus, therapy should entail both core values that underlie antisocial behaviour as well as acquiring skills to increase the individual's likelihood of remaining offence free (Ward et al., 2012).

Ward and Stewart (2004) assert that treatment goals tend to be conventionally negative, avoidant, and focus on managing risk. For instance, a goal may involve "the reduction of maladaptive behaviours, the elimination of distorted beliefs, the removal of problematic desires, and the modification of offence supportive emotions and attitudes" (Ward & Brown, 2004, p. 245). As a result, treatment goals following the RNR approach are dictated by eliminating factors instead of promoting prosocial behaviour and comprising goals that are personally satisfying. Greater treatment engagement might be expected when patients are asked to approach positive, prosocial experiences and relationships, rather than avoiding high risk situations, associates, thoughts, and emotions (Serin & Lloyd, 2009).

The GLM framework differs in this respect from other approaches. Instilling the skills, knowledge, and resources in order to assist offenders in leading different kinds of lives is imperative for rehabilitation (Ward, 2002b). Treatment according to the GLM approach includes actively working toward helping patients attain life goals (Yates, 2013). Examples of approach rather than avoidant goals include helping patients to attain independence and self-sufficiency without abusing others, and experiencing sexual pleasures in healthy ways that are not harmful to others (Yates, 2013).

The GLM has been linked to the Self-Regulation Model (SRM), a variation of relapse prevention for sexual offenders. The SRM proposes four basic pathways for sexual offending (e.g., avoidant/passive, avoidant/active, approach/automatic, and approach/explicit) (Lindsay et al., 2007). Simons, McCullar, and Tyler (2008) compared a conventional relapse prevention program to a sexual offender treatment program employing a GLM approach. Participants in the latter group displayed significantly higher engagement and levels of motivation to participate in treatment as rated by therapists. Participants in the GLM condition were more likely to complete treatment and remained in treatment longer. Further, participants in this group experienced greater improvements on social skills, victim empathy, and problem solving skills from pre to post treatment in comparison to the standard relapse prevention program. Although studies are limited in number, this study suggests that the GLM approach can be applied to individuals in conflict with the law and the results are encouraging.

Recovery. Another concept relevant to the discussion of protective factors is the recovery and resilience of mentally-disordered offenders. Recovery involves instilling hope and empowerment, encouraging autonomy, helping patients to prepare for reintegration into the community, providing good role models, and teaching ways to cope (Green, Batson, & Gudionsson, 2010). Mezey, Kavuma, Turton, Demetriou, and Wright (2010) note that symptom - reduction is necessary but not sufficient for recovery. Building a balanced and meaningful life, increasing self-confidence, and desistance from reoffending were also found to be key to the recovery process (Mezey et al., 2010). A challenge relevant to recovery among mentally disordered offenders is that individuals are legally detained and unable to control all decisions related to their care, which may lead them to feel pressured into recovery (Green et al., 2010). Moreover, recovery may be mainly guided by an attempt to manage and reduce the individual's

risk to the public, and the individual's wishes may not take priority. Additionally, it may be difficult to foster hope among patients, given that the length of institutionalization may be long, compared to general clinical patients. Nevertheless, recovery is a useful framework for understanding strength and resilience among mentally disordered offenders and for the clinical utility of protective factor instruments.

Contrasting GLM with RNR. Unfortunately, there remains a lack of empirical research on the GLM model in comparison to the widely studied RNR model (Siegert, Ward, Levack, & Mcpherson, 2007). The RNR model is a prominent and empirically-supported framework within correctional settings for guiding the actuarial assessment and treatment of offenders in Canada (see Andrews & Bonta, 1998; Andrews et al., 1990). Currently, there is ongoing debate as to whether the GLM serves to complement the RNR approach. The creators of the RNR model maintain that the lack of empirical evidence for the GLM limits its usefulness (Andrews, Bonta, & Wormith, 2011; Ward et al., 2012). In contrast, Ward and colleagues consider the GLM framework an enhancement to existing approaches, including RNR, cognitive behaviour therapy and motivational interviewing by improving treatment effectiveness and overcoming shortcomings, particularly related to poor patient compliance (Ward et al., 2012).

As its name implies, the RNR framework encompasses three central principles (Andrews et al., 2006). The Risk principle states that the risk level of an individual, as determined by evidence-based risk instruments, should correspond with the level of treatment intensity administered. Therefore, higher risk individuals should receive a higher treatment dosage, and vice versa. The Need principle states that criminogenic needs, or risk factors possessed by the individual perpetuates reoffending, and therefore should be targeted. Lastly, the Responsivity principle states that the style of the intervention employed should coincide with the learning style

and abilities of the offender. Overall, the RNR model applies to a wide range of individuals in conflict with the law, providing an all-encompassing formula for treating and assessing offenders, while also taking into account individual functioning (Andrews et al., 2006).

Both GLM and RNR models take a rehabilitative, rather than a punitive approach to working with offenders. However, their frameworks diverge with respect to their theoretical stance on human nature and therapeutic goals (Ward, 2002b). The RNR framework is criticized for being overly negative and reductionistic in its perspective of offending behaviour because its focus concerns the identification of the individual's deficits (Ward, 2002b). Strengths, values, or function of offending behaviour is not addressed in the RNR approach. Conversely, the GLM asserts that human beings are essentially good but may strive to attain their goals and values in socially unacceptable ways, thus failing to engage in prosocial behaviour (Ward & Brown, 2004). Since the GLM framework discusses the causes of human behaviour and motivation, and considered a comprehensive model, whereas the RNR model only pertains to offending behaviour (Ward et al., 2012). In general, the RNR focuses solely on reducing risk. However, the GLM is arguably a more comprehensive model by addressing both risk reduction and the promotion of goods (Ward et al., 2012).

The two approaches differ in how they conceptualize risk factors. The RNR model asserts that risk factors, also referred to as criminogenic needs, are characteristics of people and their circumstances that result in an increased likelihood of reoffending, such as an individual's peer group, substance abuse and recreational activities (Andrews et al., 2006). The GLM perspective states that risk factors can be better understood as markers indicative of problems in the way an individual seeks primary goods, due to their antisocial means of attaining them (Ward & Brown, 2004). In more simplistic terms, there may be a lack of internal or external conditions necessary

to achieve their primary goods, and risk factors may be obstacles that reduce an individual's capacity to live a fulfilling life (Ward, 2002b). Interestingly, the primary eight criminogenic needs outlined by the RNR framework (history of antisocial behaviour, antisocial personality pattern, antisocial cognition, antisocial associates, family and/or marital problems, school and/or work problems, leisure and/or recreation, and substance abuse), relate to the difficulties that individuals encounter when striving to attain primary goods (Andrews et al., 2006; Ward, 2002b). Therefore, each primary good corresponds to a risk factor identified by the RNR model.

Another noteworthy difference between the two approaches concerns the conceptualization of mental illness and its contribution towards offending behaviour. Barnao (2013) argues that one of the primary limitations of the RNR model is that mental illness is assumed to be a responsivity factor. Responsivity factors affect how an individual learns and interacts with their treatment environment, which can function to either facilitate or hinder treatment, such as an individual's motivation to engage (Andrews, Bonta, & Hoge, 1990). As such, the RNR framework does not identify mental illness as a primary contributor of crime or criminogenic need. In the case of individuals found NCRMD, this conceptualization of mental illness is inconsistent with the reason for their offending behaviour (e.g., committing an offence due to symptoms of psychosis). This point is germane to the current study because the majority of the population under investigation received a diagnosis of major mental illness and deemed NCRMD.

The GLM approach conceptualizes mental illness as either temporary or long-term, and poses a barrier to an individual's attainment of primary goods (Barnao, 2013). Further, mental illness may provide a means for which primary goods are attained. For example, violent

behaviour may be psychotically-driven, if a psychotic individual derives a sense of control from a delusional belief that he/she possess special powers (Barnao, 2013).

Overall, the GLM better accounts for individuals whose mental illness is directly related to their offending patterns and serves as a valuable framework for examining protective factors among individuals found NCRMD. Nevertheless, the GLM is best embedded within the risk needs approach and is useful for viewing patients holistically by revealing motivations for offending (Ward, 2002b). The current study does not aim to test the GLM model, but rather given that the GLM philosophy supports the identification and inclusion of protective factors in clinical services for forensic patients, it is a relevant framework worthy of review.

Not Criminally Responsible Persons

Not Criminally Responsible on account of Mental Disorder (NCRMD) is a legal defence that came into effect in 1992 in Canada for forensic patients, previously referred to as Not Guilty by Reason of Insanity (NGRI). Forensic patients are a subset of the mental health patient population who differ from correctional populations, and are under the authority of the criminal justice system (Verdun-Jones, 1994). There are no grounds to convict a person who committed an offence while suffering from a mental disorder according to Section 16 of the Criminal Code of Canada (Bill C-30, 1992). An NCRMD defence is neither a verdict of guilt nor an acquittal, resulting in either being immediately released, remaining in the community or being admitted into a psychiatric facility (Desmarais et al., 2008). Common types of symptoms found to justify an NCRMD defence include delusions, hallucinations and confusion (e.g., psychotic depression, post-traumatic stress disorder, psychosis mania and organic brain syndromes) (Bloom & Schneider, 2006).

Severe mental illness or psychosis does not necessarily equate to the absence of criminal responsibility, as these symptoms must be shown to be directly related to the index offence. The legal standard used to establish whether an individual is capable of appreciating the impact of their actions relates to the M'Naghten rule (*R. v M'Naghten*, 1843). In 1843, Daniel M'Naghten was acquitted for murdering a man, whom he mistook for the British Prime Minister (Bloom & Schneider, 2006). According to the M'Naghten rule, determining whether an individual qualifies for the NCRMD defence requires a lack of moral wrongfulness, which differs from a legal wrongfulness (Bloom & Schneider, 2006). The debate revolving around what constitutes as legal versus moral wrongfulness arose in *R. v Chaulk* (1990). Although an individual may be aware that they are breaking the law, they may not possess the insight that their behaviour is morally wrong.

Over the years, there have been two bill amendments relating to the rights and management of individuals designated NCRMD, one in 1992, the other in 1999. Prior to 1992, persons found NCR were mandated to automatic and indefinite institutionalization regardless of their risk to the public (Verdun-Jones, 1994). *R. v. Swain* played a key role in modifying the automatic and indefinite detention of people with an NCRMD status by concluding that the indefinite custody of these individuals was unconstitutional (Verdun-Jones, 1994). Specifically, the Supreme Court of Canada proclaimed in 1991 that an indeterminate detention is against Sections 7 and 9 of the Canadian Charter of Rights and Freedoms (*R v. Swain*, 1991).

In 1999, the concept of protecting society while balancing the needs of the mentally ill arose from *Winko v. British Columbia* (1999). This case clarified what is considered to be a significant threat to the public, as well as the importance of implementing the least onerous and restrictive penalty. Additionally, a significant threat signifies that a real risk of physical or

psychological harm is possible, and should be distinguished from behaviour that is merely a nuisance to others (Desmarais, Hucker, Brink, & De Freitas, 2008). Unlike the criminal justice system, the onus is on the review board to demonstrate an existing risk to support not discharging the individual (Simpson, Penney, Seto, Crocker, Nicholls, & Darby, 2014).

Once the significant threat condition is met, the least onerous and restrictive measure is ordered by the review board. The need to protect the public, the mental status of the accused, and the reintegration of the accused into the community must all be considered to inform the court or review board's decision (Bloom & Schneider, 2006). The amendment that transpired from *Winko v. British Columbia* (1999) highlights that safeguarding the rights of NCRMD persons is a priority of the criminal justice system.

Effective July 2014, the Canadian Federal Government enacted the NCRMD reform legislation, Bill C-54, as part of the tough on crime agenda. The rationale for the bill was that the public requires greater safety from mentally ill offenders, resulting in stricter conditions on NCRMD persons and holding them more accountable for their actions. Additionally, this bill allows the Crown to apply for a "high risk" designation for an accused who has caused serious personal injury to their victim(s) during their offence. Individuals with this label may receive more conditions and restrictions on their disposition, as well as have their review board hearings extended from an annual basis to every three years. Lastly, the bill allows for public and victim notification upon the discharge of NCRMD persons into the community (Canadian Bar Association, 2013). Little is known about whether these changes will result in a reduction of recidivism, and may potentially lead to increased stigmatization of mentally ill individuals (Canadian Psychiatric Association, 2013). In sum, several cases over the years have led to

Criminal Code of Canada amendments to balance the rights of the accused with the safety of the public.

The number of individuals receiving an NCRMD designation in Ontario has gradually increased over the last several years, indicating that more individuals with mental disorders are coming into conflict with the law and/or the NCRMD defense is being raised and granted more often (Ontario Review Board, 2012; Simpson et al., 2014). From 2014-2015, the courts found 71 accused unfit to stand trial and 152 were found NCRMD, totaling 223 newly accused individuals under the jurisdiction of the ORB (ORB, 2015). There are clear differences in the number of persons found NCRMD across provinces. A recent report by Crocker and colleagues (2015) found that Quebec had 6.4 times the number of cases diverted to the review board relative to Ontario, and 5 times that to British Columbia. Possible reasons for interprovincial differences and increase in NCRMD findings may be due to increased civil mental health resources and legislation changes (Crocker et al., 2015; Simpson et al., 2014).

Demographic characteristics of NCRMD persons have been examined using longitudinal research. Crocker et al., (2015b) reported figures from Ontario, and found that NCRMD persons are primarily male, approximately half have a high school diploma and over three quarters were single, at the time of the index offence. Further, approximately 65% of NCRMD persons were born in Canada and approximately three quarters of individuals were receiving some form of governmental financial assistance. Approximately 50% were living with others (spouse, family or friends) at the time of the index offence, while 20% were living alone and 10% were residing in a supervised setting. The most common index offence was assault and threatening another individual, and the victim was a stranger in 27% of cases. A psychotic spectrum disorder was the primary diagnosis in 80% of cases. The mental state of the NCRMD person at the time of the

index offence was predominantly delusional thinking (53%), followed by experiencing hallucinations (23%) and substance use and/or intoxication (22%) (Crocker et al., 2015b). Another study found similar results, reporting that hallucinations or delusions were the primary motivation for violence during the index offence among NCRMD persons (Penney, Morgan, & Simpson, 2016).

The Ontario Review Board. The Ontario Review Board (ORB) is the governing body that determines whether NCRMD patients pose a significant risk to the public. The ORB is a panel governed by federal legislation, composed of a chairperson, two psychiatrists or one psychiatrist and one psychologist, one legal member, and one public member (ORB, 2012). Members of the ORB decide upon the level of supervision necessary to manage patient risk within a hospital setting and in the community after a hearing. There are three possible outcomes: absolute discharge, a discharge subject to conditions, and detention in a hospital subject to conditions (ORB, 2012).

A host of factors are considered by review boards when formulating NCRMD dispositions. A study by Crocker and colleagues (2011) examined the factors informing decision making among review boards in Quebec. Review boards have been found to rely mainly on dynamic risk factors rather than historical static risk factors to inform their decisions. Decisions about NCRMD dispositions were only slightly associated with actuarial estimations of future risk for violence (Crocker et al., 2011). Other studies found that decisions among review boards in Ontario were highly associated with clinician recommendations, which were not always based on risk-related factors (Hilton & Simmons, 2001).

A recent study further elaborated on the relationship between recommendations made to the review board and the type of disposition assigned. Hilton, Simpson, and Ham (2016), found

that the psychiatrist's recommendations at the review board hearing was directly related to the disposition type received. Compared to findings from earlier studies, psychiatrist's recommendations were now more strongly influenced by risk assessments, specifically structured professional judgement instruments. These results suggest that that there is a trend toward the increasing influence of violence risk assessment on risk management decisions (Hilton et al., 2016). Review board decisions are also made on the basis of clinical treatment team recommendations. A recent study examined clinical treatment team recommendations to the ORB about the patient's level of security during the upcoming year (Martin & Martin, 2016). They found that recommendations for patients on a medium secure unit were based on active symptoms of psychosis and overall violence risk level, while recommendations for patients on a minimum secure unit were influenced by the number of critical incidents (e.g., elopements, late check-ins, verbal/physical aggression, substance use) occurring within the last year. Overall, it appears that review boards play an important role in determining how to best manage risk among persons found NCRMD and primarily rely on dynamic risk factors, which are intended to predict short-term changes in risk level (Crocker et al., 2011).

Risk management of NCRMD persons. Managing the potential risk of NCRMD persons is a central role of review boards, as well as the psychiatric facilities that house them. As previously noted, three disposition options are available and detail the conditions that apply in the upcoming year, based on a majority vote from panel members (ORB, 2012). An absolute discharge can be granted if the accused is not a significant threat to public safety. The accused is immediately released to the community without any conditions.

A detention order may be assigned for those deemed a significant public threat. A detention order includes detaining the accused in a psychiatric hospital, with varying levels of

privilege and security (ORB, 2012). Detention dispositions will indicate whether the accused is sent to a maximum, medium (secure) or minimum (general) security level of the hospital and mandatory conditions, with some exceptions. A conditional discharge may be granted in which the accused is approved to live in the community and must follow specific conditions such as abstaining from using drugs and alcohol (ORB, 2012). The level of supervision both within and outside the hospital is described in the disposition and unique to each person.

Rates of disposition types depend on many factors. Livingston and colleagues (2003) found that approximately one quarter of NCRMD persons spent 10 or more years under the authority of a review board. Only 2.5% of NCRMD individuals received an absolute discharge at the time of their initial hearing, and this rate remained constant over the years within Ontario (Demarais et al., 2008; Livingston et al., 2003). Also, the seriousness of the offence was found to be related to the length of time the accused was detained (Crocker et al., 2011). Compared to Quebec and British Columbia, Ontario had the longest period of institutionalization (Desmarais et al., 2008). However, this finding may be confounded by the fact that homicide was more prevalent in Ontario than the other provinces examined (Desmarais et al., 2008). In Ontario, Crocker and colleagues (2015a) found that a detention order with conditions (e.g., accompanied leave, unaccompanied leave, live in a known place, abstain from alcohol or drugs) to be the most common type of disposition. Overall, different disposition types are assigned to NCRMD persons to manage their risk in either the hospital or community.

Generations of Risk Assessment Instruments

Numerous violence risk assessment instruments have been developed over the past two decades (Heilbrun, Yasuhara, & Shah, 2010). Risk instruments provide a risk estimation in the form of a quantitative value (actuarial percentage), qualitative category (low, moderate, or high

risk category) concerning the probability that an individual will reoffend violently, sexually, or generally (e.g., less serious offences, noncontact offences), or both (Mills, Kroner, & Morgan, 2011). Forensic evaluations often focus on the deficits of the individual to make inferences and predict future offending behaviour. In addition to predicting risk of recidivism, risk assessments identify risk factors for supervision and risk management purposes (Miller, 2006).

At present, there are four generations of risk assessment instruments, all of which differ with respect to the format and risk factors included. The incorporation of protective factors in assessing risk is a novel approach and considered a 'new frontier' in forensic mental health (de Ruiter & Nicholls, 2011). The first generation of risk assessment is referred to as unstructured clinical judgement. It is an idiographic approach that relies on professional intuition to make decisions about risk, and is not an empirically supported method (Andrews et al., 2006).

Second generation risk assessment instruments use statistical methods to estimate risk. The Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993; Quinsey, Harris, Rice, & Cormier, 1998) and the Static-99 (Hanson & Thornton, 2000) are examples of second generation actuarial risk measures and incorporate a limited number of items that are primarily static in nature and have been shown to possess predictive validity (Andrews et al., 2006; Hanson & Morton-Bourgon, 2009). However, actuarial or second generation tools have limited value in a treatment setting because the incorporated risk factors are not amenable to change. Nevertheless, actuarial instruments are useful for inferring moderate to long-term risk predictions, as opposed to determining risk management and treatment planning (de Ruiter & Nicholls, 2011).

Third generation instruments demonstrate greater sensitivity to life circumstances by providing information regarding targets of treatment and revealing the potential likelihood of

recidivism (Bonta & Andrews, 2007). Third generation risk instruments, such as the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 2000) include both static and dynamic factors. This type of measure identifies eight primary factors associated with offending: history of antisocial behaviour, antisocial personality, antisocial cognition, antisocial associates, family relationships, marital relationships, school, work, leisure and recreational activities, and substance abuse (Andrews & Bonta, 2006).

Lastly, fourth generation risk assessment instruments such as the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2000), an updated version of the LSI-R, integrates actuarial measures with structured clinical judgement, outlining possible avenues for case management, and targets for intervention. Many fourth generation tools use structured professional judgement (SPJ) to establish an individual's risk level, which consists of a list of static and/or dynamic factors that are empirically validated. SPJ tools commonly provide coding guidelines to assist evaluators in exercising discretion when conducting the risk assessment (Guy, 2008). Items on the tool are interpreted and weighted by the clinician to derive a risk rating/judgement to inform future risk of recidivism for the particular individual being evaluated in a given situation (Douglas & Kropp, 2002). The HCR-20 (Webster, Eaves, Douglas, & Wintrup, 1995) and SAPROF are examples of SPJ tools.

Another characteristic of SPJ tools is that they are often constructed using input from various stakeholders, such as researchers, clinicians and administrators. An example is the development of the SAPROF, which was first created by reviewing the literature on protective factors, and then relying on the clinical expertise of mental health professionals at a psychiatric hospital. Both researchers and mental health professionals rated items based on their perceived relevance from a clinical standpoint (de Vries Robbe, 2014).

A meta-analysis by Guy (2008) examined a variety of SPJ tools across 113 studies. Most notably, the HCR-20, START, Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997) and the Spousal Assault Risk Assessment (SARA; Kropp, Hart, Webster, & Eaves, 1994) were included among the SPJ tools under investigation. The findings of the study revealed that these tools successfully predicted both violent behaviour (AUC=.74) and sexually violent behaviour (AUC=.59). The psychometric properties of individual risk and protective factor measures will be outlined later in the present study.

Risk and protective factors. Risk and protective factors are relevant to forensic assessment and the prediction of recidivism (Miller, 2006). In particular, both static/stable and dynamic/acute factors are forms of risk and protective factors. A risk factor is defined as a characteristic that precedes and increases the likelihood of an adverse outcome of interest, such as recidivism or institutional misconduct (Kazdin, 2003). Examples of static risk factors include history of violent behaviour or past substance abuse, because they are unlikely to change over time, whereas dynamic risk factors directly relate to an outcome under investigation such as recidivism, and demonstrate potential to change over time. For example, an individual's peer group, employment status, and leisure or recreational activities are considered dynamic risk factors, since they can change over time. Dynamic factors are difficult to measure because a change in the factor must directly relate to a change in the outcome, such as offending behaviour (Hanson, 2009; Mills et al., 2011). Dynamic risk factors were shown to have ecological validity from the perspective of clinicians and review board members (Crocker, Braithwaite, Cote, Nicholls, & Seto, 2011). In a recent study, dynamic risk factors were found to be more strongly associated with decisions regarding the detention or release of NCRMD persons, than historical risk factors (Crocker et al., 2011).

Another study investigated clinicians' perceptions of the factors they believed to be most influential in modifying risk of violence (Sturidsson, Haggard-Grann, Lotterberg, Dernevik, & Grann, 2004). Dynamic factors were indicated as most useful from the clinicians' perspective. Specifically, lack of insight, poor treatment motivation, receiving psychiatric treatment, contact with professional support, and substance abuse were among the most commonly reported. In general, it is evident that clinicians tend to focus on factors they consider to have an influence in changing. Nevertheless, static factors such as the offender's age and criminal history still remain the strongest indicators of future violence (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2009). Although the changeable nature of dynamic factors makes them promising targets for reducing risk, static factors are consistently the best predictors of future violence, and thus should not be overlooked during an assessment (Douglas & Skeem, 2005; Hanson & Morton-Bourgon, 2009).

A meta-analysis investigating general and violent recidivism showed that second, third and fourth generation risk instruments outperformed first generation instruments in their predictive validity (Andrews & Bonta, 2006). Further, another meta-analysis comparing the predictive ability of risk instruments for violence showed that third generation instruments, when combined with file reviews and client interviews, best predicted violent reoffending compared to second generation instruments (Campbell, French, & Gendreau, 2009). Second-generation instruments, when combined with file reviews, were strong predictors of institutional misconduct. Interestingly, theoretically-derived measures showed larger effects than measures that were not theory-based, such as the Psychopathy Checklist-Revised (PCL-R; Hare, 2003), which assesses the presence of psychopathic traits. Overall, dynamic risk factors demonstrate good predictive validity when combined with other methods.

A comprehensive model of risk assessment should include static factors to predict future risk, and dynamic risk and protective factors to inform risk management and treatment progress (de Ruiter & Nicholls, 2011). The next section will define protective factors, as well as the various models used to understand the role of protective factors in relation to recidivism.

Defining Protective Factors

Protective factors have long been a focal point of risk assessment measures designed for children and adolescents. This likely reflects the importance of resilience and strengths as developmental factors during youth, and the value of identifying protection for preventative purposes (Abidin et al., 2013; Farrington, 2000). Early definitions describe protective factors as elements that influence, modify or improve response to a person's negative internal or environmental factors, which predispose offending behaviour (Rutter, 1985). Recent conceptualizations of protective factors appear to coincide with this definition. For instance, Andrews and Bonta (2006) state that protective factors can reduce the likelihood of a maladaptive outcome and can be static or dynamic. Nevertheless, the theoretical framework underlying protective factors, and their relationship to recidivism is a complex one and frequently debated (Rogers, 2000).

The lack of agreement about the function of protective factors among researchers mainly stems from the theoretical perspective of protective factors (Ullrich & Coid, 2011). Some have argued that protective factors represent the absence of risk factors, meaning that they are mutually exclusive variables (Costa, Jessor, & Turbin, 1999). Other researchers contend that protective factors are the opposite of risk factors, suggesting that they exist on a continuum (Webster, Martin, Brink, Nicholls, & Middleton, 2004). Another perspective states that protective factors represent the opposite of risk factors for some, but not all factors (Farrington &

Loeber, 2000; Stouthamer-Loeber et al., 1993). For instance, the absence of poor social skills (a risk factor) does not necessarily equate to good social skills (a protective factor) and may be present without a corresponding risk factor. Finally, other perspectives diminish the role of protective factors in comparison to risk factors for predicting violent behaviour (Andrews et al., 2006).

The current study asserts that protective factors must characterize definable propensities or manifestations, and are not merely the absence of risk. This assertion is congruent with Rutter (1985) who stated that protective factors should represent more than simply the absence of harm or adverse outcomes. As such, it should be possible to define protective factors without using the negative form (e.g., lack of empathy). Further, both risk and protective factors can be present at the same time to varying degrees (Ullrich & Coid, 2011).

The function of protection. The ways in which protective factors operate in relation to risk and violent recidivism is complex and many models have been proposed in the literature. Given that the presence of risk factors or the absence of protective factors usually precede violent behaviour, risk factors are not mere deficits, nor are protective factors simply strengths (de Vries Robbe, 2014). Rather, a hypothesized or established timeline or temporal precedence must be indicated, in which the risk or protective factor precedes the outcome under investigation. Therefore, the presence of risk factors increases the likelihood of future violence, whereas the presence of protective factors is assumed to decrease the likelihood of future offending or increase the likelihood of desistance from future violence.

Fitzpatrick (1997) proposed two theoretical models in which protective factors serve to either mediate or moderate the relationship between risk factors and violent behaviour. The mediation model states that protective factors reduce or weaken the effect of risk on violent

behaviour. According to this model, risk factors have a direct effect on violent behaviour. Therefore, protective factors may be mediators or intervening variables that can elaborate on exactly what happens to result in the outcome, beyond knowing that risk factors cause violent behaviour (Kazdin, 2003). As a result, the indirect effects of protective factors can be accounted for through the relationship between the predictor and outcome. An example of this is the influence of medication on recidivism. Medication (an item on the SAPROF) may explain the relationship between active symptoms of major mental illness (a risk factor on the HCR-20) and desistance from reoffending.

The moderator model suggests that protective factors interact with or moderate the relationship between risk and offending. Risk factors impact violent behaviour, but only under certain conditions, such as when protective factors are absent. Accordingly, an individual may partake in physical violence when risk factors are high and protective factors are low or absent. Therefore, the combined effects of risk and protective factors are important in this model. Protective factors act as a moderator or third variable that separates risk level into subgroups, differentiating the association between protection and risk and violent offending (Baron & Kenny, 1986). To illustrate this point, the higher the risk posed by an individual, the greater the likelihood of recidivism. Therefore, only high risk offenders who possess certain protective factors on the SAPROF, such as receiving *professional care* (regular contact with mental health care professionals), may experience lower recidivism rates, compared to high risk offenders who do not receive *professional care*. A recent study supported the moderator model by demonstrating that protective factors serve to buffer risk factors on recidivism under certain conditions, in a sample of adolescents (Lodewijks, de Ruiter, & Doreleijers, 2010).

A third model regarding the function of protective factors with respect to violent behaviour is offered by de Vries Robbe (2014). The main effect model specifies that some protective factors have a direct impact on offending behaviour. When a protective factor demonstrates a main effect on desistance, it signifies that some protective factors may have an overall positive effect on individuals who refrain from offending behaviour and directly reduces the chance of violent behaviour, rather than influencing specific risk factors. Items such as *work* and *leisure activities* are examples of protective factors on the SAPROF that may have a direct effect on desistance from reoffending (de Vries Robbe, 2014).

The first two models are likely to best represent the role of protective factors in relation to risk and recidivism (de Ruiter & Nicholls, 2011; de Vries Robbe, 2014). That is, protective factors both negatively and directly affect risk by reducing or weakening the effects of risk factors and influencing the association between risk factors and offending, resulting in a compensatory effect on the relationship between risk factors and offending (Day, Wanklyn, & Yessine, 2013; de Vries Robbe & de Vogel, 2013). This conceptualization coincides with the definition of protective factors provided by the SAPROF, which defines protective factors as characteristics of a person that reduce the risk of future violent behaviours, such as an individual's attitude towards authority and motivation to attend treatment (de Vogel et al., 2012). Nevertheless, there is no agreed upon model of protective factors in the literature at present.

The current study proposes an additive or combined model of protection (Rutter, 1979; Sameroff, Bartko, Baldwin, Baldwin, & Seifer, 1998; Yoshikawa, 1994). That is, risk and protective factors are cumulative. Protective factors incrementally add to the predictive ability of risk factors for recidivism. If this is the case, it suggests that protective factors do not just represent the inverse of risk factors, but rather uniquely contribute to the prevention of violent

offending by offsetting risk (de Vries Robbe, de Vogel, Douglas, & Nijman, 2015). Thus, the identification of protective factors provides additional information regarding the relationship between risk factors and violent behaviour, over and above risk factors alone (Sutherland, Merrington, Jones, Baker, & Roberts, 2005).

Instruments Examining Protective Factors

Although there is a paucity of research on protective factors among adults, protective factors and resilience have been well researched among children and adolescents. The Structured Assessment of Violence in Youth (SAVRY; Borum, Bartel, & Forth, 2003), is the first tool designed to measure protective factors in adolescents in a correctional or forensic population. The SAVRY is an SPJ tool composed of both risk and protective items that are coded as either absent or present. There are 24 risk factor items on the SAVRY, and an additional six protective factors, namely: prosocial involvement, strong social support, strong attachment and bonds, positive attitude towards intervention and authority, strong commitment to school, resilient personality traits. The interrater reliability of the Summary Risk rating on the SAVRY was excellent (ICC=.85) (Lodewijks et al., 2008).

A recent study examined the predictive validity of the SAVRY on various disruptive behaviours in residential treatment among juveniles (Lodewijks et al., 2008). The risk items on the SAVRY predicted physical violence against other persons, violence against objects, verbal threats, and rule violations. Strong social support, strong attachments and bonds, positive attitude toward intervention, and strong commitment to school were the only protective factors that significantly predicted physical violence against other persons.

Another tool that examines risks and strengths among youth (aged 12-18) who have sexually offended is the AIM2 (The Assessment, Interpretation and Moving on Project 2;

Griffin, Beech, Print, Bradshaw, & Quayle, 2008). Four domains are assessed in the AIM2: sexually and non-sexually harmful behaviours, developmental, family/caregiver, and environment. There are 69 items; 26 are static factors, 18 are dynamic factors; 6 are static strengths and 19 are dynamic strengths, which exist across all four domains. Protective factors include use of emotional confidant, positive use of support network, good negotiation/problem solving skills, and positive relations with professionals. Griffin and colleagues (2008) conceptualize protective factors as factors that indirectly reduce trait level problems (e.g., psychological dispositions) and directly reduce triggers related to sexually abusive behaviour.

Items on the AIM2 were divided into risks and strengths scale and demonstrated acceptable reliability (Cronbach's α=0.65 and 0.69, respectively). The findings from a pilot study suggest that the risk scale and strength scale predicted sexual recidivism among a group of youth (Griffin et al., 2008). Similarly, another study demonstrated that the AIM2 predicted sexual reoffending among youth with intellectual disabilities who committed sexual offences (Griffin & Vettor, 2012).

Desistence for Adolescents who Sexually Harm (DASH-13; Worling, 2013), is a protective factor measure for adolescents who sexually offend. The DASH-13 is a checklist of 13 protective factors, scored as yes/no and totaled, ranging from 0-13. Items are divided into two domains: one domain containing factors pertaining to desistence from sexual offending, and the other domain relating to general functioning (Zeng, Chu, & Lee, 2015). Items on the DASH-13 were selected based on factors contributing to the onset rather than the maintenance of sexual offending behaviour (Zeng et al., 2015). The DASH-13 was shown to have fair interrater reliability (ICC=.54), and positively correlated with the SAPROF (Zeng et al., 2015).

The Inventory of Offender Risks, Needs and Strengths (IORNS; Miller, 2006) is another tool that incorporates protective factors, as well as offender risk and needs among adult male and female offenders. The IORNs was the first tool to comprehensively include all three factors: static risk, dynamic risk and protective factors. It is a self-report measure consisting of 130 yes/no questions. There is a total of nine factors/scales within the IORNS: static risk, criminal orientation, psychopathy, intra/interpersonal problems, aggression, alcohol/drug problems, negative social influence, personal resources, and environmental resources. The latter two scales contain protective factors. Examples of items on protective factor scales include cognitive/behavioural regulation, anger regulation, and educational training. Validity scales are included on the IORNS to assess an inconsistent responding style and favourable impression management.

The IORNS has demonstrated good internal consistency among the three indexes: static risk index (Cronbach's α=.73), dynamic risk index (Cronbach's α=.91), and protective strength index (Cronbach's α =.85). The IORNS was shown to successfully predict which offenders violated their conditions (Miller, 2006). Offenders who breached their conditions on two or more occasions were shown to have higher scores on their overall risk index, higher scores on their dynamic needs index, and lower scores on their protective strengths index, compared to offenders with one or less reoffence. The IORNS has also been normed on a sample of offenders with a low grade reading level and can be administered to offenders with various levels of education (Miller, 2006).

The Short-term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Middleton, 2004) is another SPJ tool composed of a risk/vulnerability scale and protection/strengths scale (see Measures section). The START has 20 items and each item is

coded twice, as a strength and as a vulnerability. Examples of items on the START include social skills, relationships, material resources, rule adherence, insight, and treatability.

The START has a substantial amount of research examining its psychometric properties; however the findings are mixed across published studies. The START has predicted interpersonal violence, verbal threats, and any patient aggression for the short-term only (e.g., 30-90 days) (Chu et al., 2011). A study using longer-term follow-up (e.g., 90 days) demonstrated good predictive accuracy for interpersonal violence and any patient violence, but not for verbal threats using the strength scale (Chu, Thomas, Ogloff, & Daffern, 2011). Another study found partial support for the predictive validity of the START for challenging behaviour (Braithwaite, Charette, Crocker, & Reyes, 2010). Although the START is designed to assess multiple areas of adverse events, it did not predict self-neglecting behaviour or victimization (Gray et al., 2011). Further, other studies found the strength scale to be less predictive than the risk scale when assessing future violence (Wilson, Desmarais, Nicholls, & Brink, 2010). Overall, the START appears to be most useful for estimating inpatient aggression in the short-term and is commonly compared and contrasted with the SAPROF due to its overlapping content.

The Dynamic Risk Assessment for Offender Re-entry (DRAOR; Serin, 2007), is another SPJ tool that assesses both imminent risk of reoffending and desistance from crime using protective factors. Risk factors are divided into stable risk factors (stable dynamic risk factors that are potentially changeable e.g., problem solving skills) and acute risk factors (dynamic risk factors that change rapidly e.g., substance use). The DRAOR was specifically developed to assist parole and probation officers in order to review changes in the offender's life and respond accordingly (Hanby, 2013). The DRAOR is intended to be administered to offenders by corrections staff. There are 19 items rated on a 3-point scale, ranging from not a problem/asset to

a definite problem. Items are distributed among three subscales: stable dynamic risk factor subscale, acute dynamic risk factor subscale and protective factor subscale. Items on the protective factor subscale include responsive to advice, prosocial identity, high expectations, cost/benefits, social supports and social control. Thus, protective factor items are not simply the inverse of risk factors but conceptualized as attributes that guard the individual (Hanby, 2013).

Studies examining the psychometric properties of the DRAOR revealed that the pre-test, scores were significantly correlated with any violation or return to prison (Tamatea & Wilson, 2009). Another study demonstrated that the stable dynamic risk factor and protective factor subscales had high internal reliability (Cronbach's α=.80 and .84), while the acute dynamic risk factor subscale demonstrated moderate reliability (Cronbach's α =.62) (Hanby, 2013). The DRAOR predicted recidivism upon a 2-year follow-up for the stable and acute subscales and protective factors subscale. The findings also revealed that at follow-up, recidivists had higher stable risk and lower protective scores compared to the pre-test. Further the DRAOR demonstrated incremental predictive validity over and above the RoC*RoI, a measure developed in New Zealand, which assesses static risk of reconviction and imprisonment (Hanby, 2013). This suggests that the DRAOR can predict reoffending over and above a risk measure alone.

SAPROF. Since its development in 2011, there have been 13 published studies examining the psychometric properties of the SAPROF (Abidin et al., 2013; Coid, Kallis, Doyle, Shaw, & Ullrich, 2015; Davoren et al., 2013; de Vries Robbe, de Vogel, & de Spa, 2011; de Vries Robbe, de Vogel, & Douglas, 2013; de Vries Robbe, de Vogel, Douglas, & Nijman, 2015; de Vries Robbe, de Vogel, Koster, & Bogaerts, 2015; de Vries Robbe, de Vogel, Wever, Douglas, & Nijman, 2016; Doyle et al., 2014; Klein et al., 2012; Turner et al., 2014; Yoon, Spehr & Briken,

2011; Zeng, Chu, & Lee, 2015). The findings across studies are generally consistent and demonstrate the utility of the SAPROF as a clinical measure.

The changeability of SAPROF scores has been used to identify treatment progress among forensic patients. de Vries Robbe and colleagues (2011) examined whether SAPROF scores improved as patients progressed through treatment. A total of 126 forensic patients receiving treatment at a psychiatric hospital in the Netherlands were included. Repeated assessments of each patient showed significant improvements of SAPROF scores over the course of treatment, at 1, 2 and 3 year follow-ups. Pre-and post-treatment ratings revealed significant positive change for dynamic factors on the SAPROF (e.g., items 3-14). Further, a significant decrease in scores were demonstrated for dynamic factors, which are expected to diminish over time with intervention (e.g., items 15 to 17).

Analyses also demonstrated good interrater reliability and good predictive validity for desistence from violent reoffending after treatment. SAPROF total scores, SAPROF Final Protection Judgements (i.e., the level of available protection for relapse into violence, composed by interpreting, weighing and integrating present protective factors) and HCR-20-SAPROF index (a corrected risk score computed by subtracting SAPROF total scores from HCR-20 total scores) predicted violence better than HCR-20 total scores at all follow-up periods. Overall, these results demonstrate that the addition of protective factors contributes information about violent reoffending over and above risk factors alone. Also, given that the SAPROF was shown to be sensitive to change over the course of repeated administration, it can be an effective tool for guiding treatment planning and evaluation.

Another study assessed the changeability of SAPROF scores during treatment to predict institutional misconduct. de Vries Robbe and colleagues (2016) looked at changes in static and

dynamic factors across different stages of treatment, as well as the predictive validity of the SAPROF and HCR-20 for aggressive incidents during hospitalization. A total of 185 patients in a forensic hospital in the Netherlands participated in the study. Participants comprised both male and female patients who committed violent and sexual offences. Patients were sentenced to mandatory treatment consisting of four stages: intramural treatment without leaves, intramural treatment with supervised leaves to the community, intramural treatment with unsupervised leaves to the community, and transmural treatment where patients reside in private or hospital housing within the community while remaining supervised by a hospital community treatment team. Incidents of aggression were defined as physical or threatening verbal aggression, occurring at least 10 months following treatment.

As expected, HCR-20 scores were lower for patients in later stages of treatment, whereas SAPROF scores were higher during later stages of treatment. Also the HCR-SAPROF index was more highly correlated with incidents of violence, than HCR total scores or SAPROF total scores individually. Overall, historical items on the HCR-20 were most predictive of violent incidents during treatment while motivational dynamic factors on the SAPROF were most predictive of inpatient aggression at 12-month follow-up. The best predictors on the SAPROF were *coping, self-control, work,* and *attitudes toward authority*. Other studies using the SAPROF also found these SAPROF items to be most the predictive of violence (Abidin et al., 2013; de Vogel & de Ruiter, 2006). The results of de Vries Robbe and colleagues' (2016) study demonstrate that protective factors are higher at later stages of treatment and that patients displaying the greatest change during treatment showed significantly lower rates of aggressive incidents after the completion of treatment. As a result, the SAPROF predicted inpatient aggression, which may inform treatment progress and treatment planning.

Previous studies explored risk management issues, specifically the level of security assigned to forensic patients. Davoren and colleagues (2013) examined whether forensic patients detained in a secure unit could be distinguished from those who were discharged by the review board. A total of 56 patients residing at a forensic psychiatric hospital in Ireland were included. Several measures were administered, including: The Dangerousness Understanding, Recovery and Urgency Manual-3 and -4 (DUNDRUM-3 and DUNDRUM-4; Flynn, O'Neill, McInerney, & Kennedy, 2011), a measure of progress in treatment and recovery, HCR-20, START, SAPROF, GAF, Suicide Risk Assessment and Management Manual (S-RAMM; Bouch & Marshall, 2003), a suicide risk assessment and management manual, and the Positive and Negative Syndrome Scale (PANSS; Kay, Fiszbein, & Opler, 1987), to assess symptoms of schizophrenia.

All measures, with the exception of the S-RAMM, successfully differentiated which patients received a discharge from the hospital. The SAPROF significantly predicted which patients received a conditional discharge, and the most predictive items on the SAPROF included *coping*, *self-control*, *financial management*, *life goals*, and *social network*. The GAF was the strongest predictor of conditional discharge compared to all other measures. The findings illustrate that SAPROF items are not equally influential; some had larger effects than others. This suggests that specific items may be an important focus for assessment and treatment.

Another study also found specific items on the SAPROF to be more predictive of violence than others. Coid et al., (2015) examined predictive and causal models of SPJ tools, specifically the SAPROF and HCR-20 V3. A total of 409 patients discharged from a forensic hospital in England and Wales participated in the study. Scores on measures were collected at pre-discharge, and 6 and 12-months post discharge to predict violence. *Self-control* demonstrated

an independent protective effect for predicting desistance from violence in the causal model. Clinical items on the HCR-20 V3 were most discriminative and associated with violence in the causal model. Thus, *self-control* was found to be an explanatory factor for violence in forensic patients, whereas other items on the SAPROF did not discriminate which patients were not violent over a 12-month period. Thus, identifying individual protective factors may inform clinicians of the factors serving to mitigate risk.

Doyle and colleagues (2014) examined the characteristics of individuals discharged from a forensic hospital and prison. A total of 568 individuals from hospitals and prisons in England and Wales were included. The HCR-20, SAPROF and other measures were collected. Analyses revealed that individuals who were discharged into the community had higher SAPROF scores than inpatients/residents. Also, individuals discharged from prison were less likely to have protective motivational factors (*motivation for treatment* and positive *attitudes towards authority*). This suggests that individuals who are discharged into the community tend to have more protective factors on the SAPROF than individuals who remain institutionalized, with apparent differences between prison and hospital samples.

The prediction of violence toward others and self-inflicted violence, was another outcome predicted by the SAPROF. In a prospective study conducted by Abidin and colleagues (2013), risk factors, protective factors and psychological tests were used to predict violence toward others and self. A total of 98 forensic patients from a psychiatric hospital in Ireland participated in the study. Several measures were administered to assess violence, defined as any actual, attempted or threatened harm to self or others. Specifically, the HCR-20, DUNDRUM-3, DUNDRUM-4, START, SAPROF, GAF, S-RAMM, and PANSS were administered.

Results indicated good internal consistency across the 17 items of the SAPROF. Both the SAPROF and START strongly and significantly correlated with one another (r = .81), indicating that they measure overlapping constructs. Also, a strong inverse correlation between the SAPROF and the clinical items on the HCR-20 was found, indicating that these measures may represent different directions on a dimension. At follow-up, 20 of the 98 participants engaged in a violent incident directed towards others or self. The SAPROF predicted the absence of both violence and self-harm. Interestingly, the SAPROF predicted self-harm almost as well as the S-RAMM. Significant main effects and an interaction effect for SAPROF scores and dynamic factors on the HCR-20 were found, revealing that protective factors can serve to offset risk when comparing violent to non-violent individuals. Also, each item on the SAPROF was a predictor of violence and self-harm, performing better than the START. Overall, the implications of this study are that tools designed to measure protective factors can complement existing risk assessment and psychological instruments, rendering them useful for treatment planning and risk management.

Previous studies examined treatment progress for predicting sexual and violent recidivism using the SAPROF. de Vries Robbe et al., (2015) investigated whether changes in dynamic factors (e.g., increases on the SAPROF and decreases on the HCR-20) at pre and post-intervention predicted violent recidivism at follow-up. A total of 108 male patients with a history of sexual and violent offending participated in the study. Participants took part in mandatory treatment that used a holistic treatment approach employing cognitive behavioural and relapse prevention strategies to reduce violence risk. The HCR-20 and SAPROF were scored at pre- and post-treatment using file information. Participants were assessed to determine whether scores on dynamic factors predicted violent recidivism at a 12-month follow-up. Change scores were

calculated by subtracting pre-treatment scores from post-treatment scores for all subscales of the SAPROF and HCR-20.

The results revealed that although risk and protective factor scores did not differ at pre-treatment between groups, recidivists had higher scores on the HCR-20 and lower scores on the SAPROF at post-treatment, compared to those who desisted from crime over the long-term. Therefore, non-recidivists improved significantly more during treatment, in terms of increasing protective factors. Internal and motivational subscales on the SAPROF appeared to account for increased change, whereas clinical scales on the HCR-20 appeared to account for decreased change. Additionally, change scores were examined at 1 year and long-term follow-up (based on each individual's maximum follow-up time available). Change scores for the HCR-20 were not significantly predictive of violent recidivism at 1-year follow-up but were significantly predictive for long-term follow-up. Change scores for the SAPROF were significant predictors at both 1 year and long-term follow-up.

de Vries Robbe et al., (2013) also examined the predictive validity of the SAPROF among forensic psychiatric patients who committed a violent and/or sexual offence. A total of 188 male patients were included in the study. The SAPROF demonstrated good predictive validity at 1 year, 3 years, and long-term follow-up.

Further, dynamic factors on the SAPROF were found to be the strongest predictors of desistance from violence, even at long-term follow-up, compared to historical items on the HCR-20. Incremental predictive validity was also demonstrated using the HCR-SAPROF index and predicted violent recidivism significantly better than the HCR-20 alone at long-term follow-up.

Irrespective of the patient's risk level, those with at least a moderate level of protection (based on Final Protection Judgement Ratings) were 10 times less likely to reoffend within the

first year after discharge compared to those with low levels of protection (de Vries Robbe et al., 2013). Among offenders with moderate to high risk scores, the presence of protective factors appeared to account for rates of reoffending. These findings suggest that dynamic protective factors are valuable in predicting desistance from violence in the long-term.

The predictive validity of the SAPROF for sexual reoffending was assessed across recent studies. de Vries Robbe and colleagues (2015) examined risk of future sexual and nonsexual violence among 83 male sexual offenders discharged from Dutch forensic hospitals. Follow-up took place at 1 and 3 years after discharge, as well as over the long-term. The SAPROF, HCR-20 and SVR-20 (an SPJ risk tool used to assess sexual recidivism) were administered. Also, Index Scores were computed (HCR-SAPROF index and SVR-SAPROF index).

The SAPROF total score was strongly negatively correlated with the HCR-20 total score (r= -.83) and the SVR-20 total score (r = -.39). The SAPROF showed good predictive validity for violent recidivism at 1 year, 3 years, and long-term follow-up. Predictive validity for sexual recidivism at 3 years and long-term follow up was also good. *Coping, self-control, motivation for treatment and attitudes toward authority, leisure activities, professional care* and *external control* all significantly predicted general violence. The HCR-SAPROF index demonstrated significantly better predictive validity than the HCR-20 total score for violent recidivism at long-term follow-up. Also, the SVR-SAPROF index predicted future violence significantly better than the SVR-20 at 1 year and long-term follow-up. These results indicate that SAPROF total scores and Index Scores can inform both violent and sexual recidivism over the long-term for individuals with sexual offending histories

Given the effectiveness of the SAPROF in predicting sexual reoffending, the type of offence committed was analyzed as a potential moderating factor. de Vries Robbe and colleagues

(2013) examined whether a violent or sexual offence served to moderate the relationship between risk and protection on recidivism. Patients were followed-up at 1 year, 3 years, and on a long-term basis subsequent to discharge from a psychiatric hospital. Both the SAPROF and HCR-20 were coded retrospectively from patient's hospital files. Results did not reveal a moderating effect for the type of offence, signifying that risk and protective factors may operate in similar ways regardless of whether a violent or sexual offence was committed.

While previous studies examined protective factors among forensic inpatients, Yoon and colleagues (2011) investigated protective factors in a correctional outpatient setting with sexual offenders. A total of 30 male sexual offenders on probation or parole participated in this pilot study. The SAPROF, SVR-20, Static-99 and PCL-R were all administered. The SAPROF was negatively correlated with the SVR-20, but not the Static-99. Negative attitude toward intervention on the SVR-20 was significantly correlated with *coping, motivation for treatment, attitudes toward authority* and *professional care* items on the SAPROF. Additionally, *self-control* was significantly negatively correlated to past non-sexual violent offences and physical harm to victim(s) in sex offence items on the SVR-20.

The findings from this study have implications for the type of instruments used in conjunction with the SAPROF. The SAPROF may better correspond with instruments comprising dynamic items rather than static, as demonstrated by its relationship to the SVR-20, but not the Static-99. Also, when individual factors are analyzed, factor combinations may be apparent and should be considered for intervention with offender populations.

Similarly, Turner and colleagues (2014) examined a subgroup of sexual offenders in relation to SAPROF scores. A total of 246 incarcerated sexual offenders who offended against children in Austria were included in the study. Sexual offenders were categorized according to

their relationship to the victim (those that worked with the victim, intra-familial offenders and extra-familial offenders). Several risk measures were administered in addition to the SAPROF. Individuals that worked with the child victim had the highest SAPROF scores. Further, the SAPROF predicted desistance from any recidivism in all child sexual abusers. Thus, the SAPROF predicted reoffending among subgroups of sexual offenders.

The SAPROF has also been applied to youth. A pilot study by Klein and colleagues (2012) examined the relationship between protective factors and risk factors among male juvenile offenders. Sixty-six juveniles who committed sexual offences were administered the STAYSOR (Screening Tool for the Assessment of Young Sexual Offenders' Risk), comprising select items from the STATIC-99, SAVRY and SAPROF. In addition, two measures were administered to assess for psychopathology. The results demonstrated that measures assessing psychopathology were positively correlated with risk factors and negatively correlated with protective factors. Further, the SAPROF total score and SAPROF Final Protection Judgement were positively correlated with the protective factors scale of the SAVRY (r =0.80) and (r =0.79), respectively. Overall, risk factors were found to be significantly negatively correlated with protective factors scores and final judgements among juveniles.

Other studies examining the psychometric properties of the SAPROF using youth were recently conducted. Desistence from sexual reoffending was examined among adolescent male sex offenders receiving probation in Singapore (Zeng et al., 2015). The study used a retrospective design and measures were scored based on file information. The SAPROF and the DASH-13 (Worling, 2013; a protective factor tool assessing adolescent sexual offending) were found to be inversely related to the ERASOR (Estimate of Risk of Adolescent Sexual Offence Recidivism; Worling & Curwen, 2001), a short-term risk measure assessing adolescent sexual

offending. Neither the SAPROF nor the DASH-13 predicted desistence from sexual or nonsexual reoffending. Although the SAPROF did not significantly predict desistence from reoffending, it was better at predicting desistance from nonsexual violence than sexual desistence. These findings suggest that more protective factors are not necessarily predictive of lower incidences of sexual recidivism among adolescent sex offenders, as was demonstrated among adult sexual offenders.

Overall the increased volume of studies examining protection is indicative of the growing interest in integrating protective factors and patient strengths in forensic assessment. The SAPROF is a novel tool that has demonstrated good reliability and predictive validity for sexual recidivism, violent recidivism, and self- inflicted violence, over short and long-term follow-ups. Also changes in dynamic factors on the SAPROF were predictive of recidivism and institutional misconduct after treatment and during later stages of treatment. Further, specific items on the SAPROF predicted changes in the level of security assigned by a review board. The SAPROF has recently been applied to various populations including forensic patients, correctional populations, inpatients, outpatients, males, females, and youth. Although these studies have contributed to the generalizability of the SAPROF, further replication of these findings is still required.

Study Rationale

Protective factors in a forensic context. The justification for the inclusion of protective factors in a forensic context includes building on strengths, increasing patient motivation, preventing the overclassification of risk level, informing risk management practices, and developing targets for intervention. As previously mentioned, the GLM model emphasizes the patient's values and helps to rebuild an identity and path towards a prosocial lifestyle. As such,

the inclusion of protective factors in clinical services is aligned with the GLM framework. An emphasis on deficits can be stigmatizing and creates a sense of injustice among forensic patients, due to the oversight of progress and accomplishments (Attrill & Liell, 2007). Moreover, strictly attending to risk factors may convey that the patient's risk factors are chronic and irreversible (Rogers, 2000). Overall, the inclusion of protective factors allows for greater consideration of patients' competencies and capabilities as opposed to solely concentrating on deficits. Further, increased emphasis on protective factors during assessment and treatment may potentially enhance insight into the recovery process, helping to foster motivation (de Ruiter & Nicholls, 2011). Thus, if patients feel as though they are being viewed holistically by acknowledging their strengths, they may show a greater readiness to engage in treatment (Ward & Brown, 2004).

The neglect of protective factors in forensic risk assessments raises concerns. Thorough and comprehensive assessments in the field of psychology require examining multiple aspects of an individual's functioning to ensure predictions are accurate. As such, risk instruments that comprise only risk factors are imbalanced, fundamentally inaccurate and limit the prediction of recidivism (de Ruiter & Nichols, 2011; Rogers, 2000). The overprediction of risk level is another potential danger of excluding protective factors from forensic assessments. Assessment measures primarily focus on negative aspects related to reoffending, as opposed to positive aspects related to desistance from offending, creating the possibility for highlighting undesirable characteristics (Miller, 2006; Rogers, 2000). As a result, assessments that focus only on risk may inflate the chance of false positives, falsely predicting that an individual will re-offend when, in actuality, they will not (Rogers, 2000).

Identifying base rates of the outcome under investigation is an important concept worthy of discussion. Violent and sexual recidivism are common outcome variables examined in

research studies. Although essential for ensuring the safety of the public, the occurrence of such an event is arguably too rare to successfully predict without undue error. Even the most accurate test will produce false positives when examining infrequent or uncommon conditions (Craig, Browne, Stringer, & Beech, 2004). Additionally, base rates can fluctuate according to which definition of violent offending is applied and the type of data used to monitor the information (e.g., charges, arrest, conviction, behaviours) (Craig et al., 2004).

The overestimation of recidivism is especially problematic, as some studies report a relatively low incidence of violence among forensic patients and rates tend to vary (Crocker et al., 2011; de Ruiter & Nicholls, 2011). A review of the base rates of violent behaviour among civilly committed individuals widely ranging from 7.5% to 66.7% (Hiday, 1990). Further, the level of risk tends to fluctuate over time and even high-risk individuals only demonstrate violent behaviour a small percentage of the time under certain circumstances (de Ruiter & Nicholls, 2011). Therefore, it is difficult to predict rates of violence among NCRMD patients because it is a low base rate behaviour.

Overclassifying the dangerousness of forensic patients can be costly for both the offender and public (Ullrich & Coid, 2011). For the patient, this may result in an increased likelihood of adverse consequences, including greater restrictions in the hospital or limited access to the community. Costs to the public relate to significant financial support required to manage risks such as probation, parole, and hospitalization. Also, assigning a higher level of risk than warranted may result in the inappropriate allocation of resources such as increased treatment dosage for cases that do not require greater treatment intensity, which has been shown to have detrimental effects (Andrews & Dowden, 2006; Bonta, Wallace-Capretta & Rooney, 2000). Similarly, overclassifying risk level can result in low-risk offenders receiving extensive and

prolonged intervention, violating the risk principle of the RNR model (Andrews et al., 1990; Craig et al., 2004). Hence, unnecessarily restricting personal liberties and misusing limited resources is a potential consequence of neglecting protective factors (de Ruiter & Nicholls, 2011).

Limitations of the literature. Previous studies examining the utility of the SAPROF have been limited due to the population included. Although the SAPROF has been used with various populations, such as male inpatients, past studies have primarily comprised Dutch, Irish, English and Singaporean forensic patients. Additionally, outpatients and female patients have occasionally been excluded from the sample under investigation, further limiting the generalizability of the findings reported. Thus, this dissertation will extend previous works by providing an account of Canadian forensic patients, including males, females, inpatients and outpatients.

Further, due to differences in criminal justice legislation across countries, the samples used have varied in their legal designation. Therefore, the current study is unique, as it is the first to use a Canadian sample governed by a Canadian review board. To further illustrate this point, relevant past studies have incorporated forensic patients with a tbs-order ('terbeschikkingtelling'), a designation assigned in the Netherlands. A tbs-order requires dangerously violent offenders to receive treatment in a hospital setting. These individuals are considered to have severe psychopathology and are held criminally responsible for their actions. As a result, most patients possess a primary diagnosis of a personality disorder and/or substance use disorder. In contrast, participants in the present study all received an NCRMD designation and are therefore *not* criminally responsible for their offence due to their mental illness. While personality and substance use disorders are commonly diagnosed among NCRMD persons, it is

typically comorbid with a primary diagnosis of schizophrenia, given that psychotic symptoms are usually the basis for committing the offence. Therefore, the predictive validity of the SAPROF may vary depending on the symptom presentation, and sample examined.

Also, past studies validating the SAPROF scored the measure using only patient files containing biographical information, psychological reports, and court reports to code items (de Vries Robbe et al., 2011; de Vries Robbe et al., 2013; de Vries Robbe et al., 2015a; de Vries Robbe et al., 2015b). The current study is the first known to employ an in-person interview, in addition to collecting file information to code items on the SAPROF. As such, incorporating self-report interviews provides more information to score items which cannot be found in the patient's file. Given that the SAPROF was designed to supplement HCR-20 scores, it is ideal to use equivalent methods, which is another advantage of conducting semi-structured interviews to score the SAPROF. The benefits of scoring the SAPROF in a similar setting to how it would naturally be administered, rather than in a highly controlled environment, is that it allows for greater ecological validity and generalizability of the findings (Penney et al., 2014).

Since one of the primary objectives of risk assessment is to inform risk management practices, the outcome variables included in the current study focused on the management of risk, rather than the prediction of long-term reoffending. The present study investigated changes in the level of security outlined in patient disposition as granted by the ORB, whereas other studies mainly focused on short and long-term rates of recidivism. Changes in security level are of particular importance considering that decisions implemented by the ORB have a direct impact on where the patient should reside, as well as the necessary conditions for effectively managing their risk. In essence, risk management practices impact both the well-being and safety of the patient, co-patients, treatment staff and the public, having both immediate and far-reaching

effects. Moreover, attrition rates in studies can be high due to long follow-up times, and the underreporting of crime can underestimate true offence rates (MacDonald, 2002). Therefore, although recidivism may be useful for exploring the predictive ability of risk or protective factor instruments, recidivism rates may not always be accurate or relevant to managing risk in the short-term.

Other outcomes examined include the administration of PRN medication, institutional misconduct and disposition breaches. Treatment staff may prescribe PRN medication when patients become agitated, anxious or aggressive towards other patients and staff (Douglas, Guy, & Hart, 2009; Quanbeck, 2006). Additionally, institutional misconduct and breaches of disposition conditions may be common precursors to violent behaviour following patients' release into the community, making them variables worthy of further exploration (Trulson, DeLisi, & Marquart, 2011).

In summary, the current study addresses these gaps in previous research and provides new information regarding the utility of the SAPROF in three ways. First, the study sample included Canadian forensic mental health patients. Second, the sample differed from other studies since it primarily included patients with a psychotic disorder who received an NCRMD defence. Third, the outcome variables examined in the present study (e.g., proxies/indicators of recidivism and risk management decisions) are informative and have rarely been the focus of prior studies.

Current Study

The current study advanced knowledge of risk assessment by examining the psychometric properties of the SAPROF, a novel measure consisting of protective factors. Various forms of reliability and validity were investigated in a sample of forensic inpatients and outpatients at Ontario Shores Centre for Mental Health Sciences, a forensic mental health hospital, in Whitby Ontario. The study was conceived and developed in collaboration with clinical forensic staff at Ontario Shores. Ensuring the use of empirically supported measures that meet scientifically rigorous standards for conducting risk evaluations is a priority at Ontario Shores.

The methodology implemented was the first to use semi-structured interviews to score the SAPROF, adding to the ecological validity of the study. The reliability and validity of the SAPROF was investigated by determining its association with risk, protection, mental status and general functioning measures. Further, the predictive validity and incremental predictive validity of the SAPROF was examined across outcomes related to patient behaviour and review board decisions, over two time points. One of the primary purposes of the study was to determine whether the addition of the SAPROF increased the accuracy of the violence risk assessment.

Research Questions and Hypotheses

1. Is the SAPROF positively correlated with other protective factor measures and negatively correlated with risk factor measures?

   Hypothesis: Convergent validity was expected between the SAPROF and other risk and protection measures. In particular, the SAPROF was expected to be positively correlated with the START Strength Scale and inversely correlated with the HCR-20, LS/CMI, and PCL-R, given that risk is a related construct to protection.

2. Is there little or no relationship between the SAPROF and general functioning and mental status measures?

   Hypothesis: Divergent validity was anticipated between the SAPROF, general functioning and mental status measures. Specifically, little or no association between the SAPROF, GAF and SANS/SAPS was anticipated, since that the SAPROF is representative of protective factors, as opposed to mental health symptoms and functioning.

3. Do SAPROF scores predict the following proxies of recidivism: PRN administrations, institutional misconducts and breaches of disposition over the past year and at a six-month follow-up?

   Hypothesis: It was anticipated that higher scores on the SAPROF would predict the absence and decrease of PRN administrations, institutional misconduct and disposition breaches over both time points.

4. Can SAPROF scores predict risk management decisions e.g., changes in privilege level and security level? In particular, do SAPROF scores predict changes in security levels between disposition types (medium, minimum, conditional discharge, absolute discharge) and privilege level within the hospital?

   Hypothesis: It was expected that SAPROF scores would predict an increase in privilege level and decrease in supervision level, which are indicative of fewer restrictions and greater freedoms.

5. Does the SAPROF provide incremental predictive validity for proxies of recidivism and risk management decisions, over and above HCR-20 V3 alone? Is incremental predictive validity established over the previous year and upon the six-month follow-up?

   Hypothesis: It was anticipated that the SAPROF would add incremental predictive validity, over and above the HCR-20 V3 to predict the presence of PRN administrations, institutional misconduct, disposition breaches, increase in privilege level and decrease in security level over the past year and upon a six-month follow-up. Further, Index Scores (e.g., corrected risk scores, calculated by subtracting SAPROF total scores from HCR-20 V3 total scores) were expected to be more predictive of these outcomes compared to the HCR-20 V3 or SAPROF alone.

6. Does the relationship between SAPROF scores, proxies of recidivism and risk management decisions differ based on patient characteristics? Specifically, do SAPROF scores differ on outcome variables based on the type of patient (inpatient vs. outpatient), length of patient institutionalization (in years), length of time on current disposition (in years), severity of the

index offence (severe/contact offence or not), number of psychological treatment sessions attended (individual and group), and participation in vocational activities (volunteer inside and/or outside of the hospital)?

Hypothesis: Patient characteristics were not expected to influence the relationship between SAPROF scores and the outcomes of interest. Therefore, the main findings should remain robust and generalizable across these relevant characteristics.

Methods

Participants

Fifty patients with an NCRMD status from the Ontario Shores Centre for Mental Health Sciences comprised the study sample. All inpatients and outpatients residing at the facility deemed NCRMD were eligible for inclusion in the current study. Adult male and female patients were included in the study. Given that the forensic program at Ontario Shores comprises patients with a diversity of mental health issues (e.g., psychosis, personality disorders, substance abuse), eligible patients had a range of diagnoses and were not excluded on that basis of any particular mental illness. The only exclusion criteria for the study were if the patient was incapable of providing consent and if their primary language was not English, as the study required participation in an interview with the researcher who is English- speaking.

Capacity to consent. Patients deemed capable to provide informed consent were eligible to participate in the study based on a two-stage assessment of the capacity. Nine clinical forensic psychologists used their clinical judgement regarding the patient's capacity based on their presentation during the informed consent process of the risk assessment interview that each psychologist conducted. As a rule of thumb, patients who were deemed incapable of providing consent for the risk assessment interview, were also determined incapable to provide consent for the current study and excluded at Stage 1. Patients interested in participating in the study and found capable to provide informed consent had their information relayed to the researcher who contacted the patient, and capacity was re-assessed at Stage 2.

The second stage of assessing capacity occurred after the consent form was reviewed between the participant and the researcher, see Appendix A. Participants were notified that they could ask questions throughout the informed consent process. The researcher asked each

participant four questions to ensure they understood the study and what their participation entailed. The following questions were posed; 1) In your own words, please describe what the study is about; 2) What are some benefits of participating in this study?; 3) What are some risk of participating in this study?; and 4) Will your participation influence decisions made by the ORB? These questions were based on the factors to consider for obtaining consent, outlined by the Personal Health Information Protection Act (PHIPA, 2004).

Patients were permitted to refer to the consent form for assistance. If the patient was unable to answer the first three questions, they were informed of the correct answer and asked again until they sufficiently demonstrated that they understood. The fourth question was not mandatory but believed to be important in order to ensure that the patient was not participating out of obligation.

Patients with a substitute decision maker (SDM) for treatment purposes only (e.g., to receive antipsychotic medication), were eligible to participate in the study. These patients still possess the capacity to consent to a research interview or clinical assessment. Generally, the threshold for providing consent is lower for an assessment or research than treatment. Since forensic clinical psychologists at Ontario Shores were the first point of contact for patients, they eliminated patients requiring an SDM for all services/assessment, as these patients were unable to consent to the risk assessment interview either.

There was no opportunity to access the patient's file for their SDM status after consent was provided, before the interview commenced. Also, the clinical forensic psychologists did not have access to this information at the time of recruitment because it is not referred to by psychologists prior to the risk assessment interview. Moreover, once a patient expressed an interest in participating in the current study, the consent process and study usually commenced

immediately. Therefore, whether the patient had an SDM was not as relevant to the recruitment process.

Setting. Ontario Shores Centre for Mental Health Sciences is a general psychiatric hospital in Whitby, Ontario. The hospital houses six inpatient forensic units, approximately 138 beds (20-26 beds per unit) and a forensic outpatient service. There are approximately 250 forensic inpatients and outpatients. Three of the units are medium secure: Forensic Assessment Unit (FAU), Forensic Assessment and Rehabilitation Unit (FARU), and Forensic Rehabilitation Unit (FRU). The remaining three are minimum secure (general): Forensic Psychiatric Rehabilitation Unit (FPRU), Forensic Transitional Unit (FTU), and Forensic Community Reintegration Unit (FCRU). The Forensic Outpatient Service (FOS) provides community-based rehabilitation to patients discharged from the hospital and remain under the jurisdiction of the ORB.

Measures

SAPROF. The Structured Assessment of Protective Factors for Violence Risk-2nd edition (SAPROF; de Vogel et al., 2009) is a relatively new SPJ tool developed to examine protective factors among adults with a mental disorder and a history of violence (de Vries Robbe, 2014). The SAPROF comprises 17 items rated on a 3-point scale. A score of 0 indicates that the item/protective factor does not apply based on the information available; 1 indicates that the item probably or partially applies; and 2 indicates that the item definitely applies (de Vogel et al., 2011). Items 1 and 2 are static factors, whereas the remaining items are dynamic. Items 3 through 14 are expected to improve during treatment, since they concern internal motivation and social functioning. Items 15 through 17 are expected to decrease/reduce with treatment because they concern the external supports necessary to manage risk (de Vries Robbe et al., 2011). As

well, items on the SAPROF are divided into three subscales: Internal subscale (e.g., *intelligence, secure attachment in childhood, empathy, coping, self-control*); Motivational subscale (e.g., *work, leisure activities, financial management, motivation for treatment, attitudes towards authority, life goals, medication*); and External subscale (e.g., *social network, intimate relationship, professional care, living circumstances, external control*).

The developers of the SAPROF used three methods of test development. First, a rational-theoretical-approach was used to validate the SAPROF, by selecting test items on the basis of a literature review on protective factors. Next, during case conferences that took place at a psychiatric hospital, the researchers asked the attending mental health professionals to rate patients on the protective factor checklist devised from the literature review (de Vogel et al., 2012). Staff ratings were based on the how applicable the items were for risk management and treatment, and how relevant the items were to their daily work and clinical experience (de Vogel et al., 2012). Next, an empirical approach was implemented by establishing criterion-related validity by ensuring that items on the SAPROF predict a criterion, such as sexual and violent recidivism (Gregoire & Jungers, 2007). The internal consistency approach of the SAPROF was also established to ensure that the items load together to support the construct being measured by the test (Gregoire & Jungers, 2007). A factor analysis was conducted revealing a multidimensional scale. Any items that did not have good internal consistency or were difficult to code were eliminated (de Vogel et al., 2012).

The SAPROF also allows users to indicate key and goal factors on the measure. Key factors are items identified as being important protective factors at the time of the assessment and signify strong protectors against violence risk. Goal factors are items identified as possessing good targets for treatment. Therefore, items on the SAPROF can be differentiated on the basis of

having a protective effect (key factors) and having the potential to possess a protective effect following intervention (goal factors) (de Vogel et al., 2011).

For the present study, items on the SAPROF were scored based on information gathered during the research interview and file information. Specifically, items on the SAPROF such as *empathy, self-control* and *coping skills*, all required more context or insight compared to other items and were corroborated with file information.

Information from the previous year was used to score items on the SAPROF, to predict protection over the upcoming 6-12 months (de Vries Robbe, 2014). The SAPROF is intended to predict desistence from offending in the long-term (e.g., six-months or greater), compared to other measures such as the START which are more short-term in scope (Webster et al, 2004).

The developers of the SAPROF designed the tool with the intention that it be used in conjunction with risk scores from other risk assessment measures, specifically the HCR-20 (Webster et al., 1997). Higher scores on the SAPROF are indicative of the presence of more protective factors, while lower scores on the SAPROF are indicative of the presence of fewer protective factors. One exception is that higher scores on items 15 to 17 represent the presence of greater external control needed to manage the individual's risk and that the individual is likely to possess lower levels of overall protection. Conversely, individuals with lower scores on items 15 to 17 will likely possess higher levels of overall protection because less external control is required to manage their risk.

Once the SAPROF is administered and items coded, a Final Protection Judgement is determined, which is an overall categorical rating of low, low-moderate, moderate, moderate-high or high. The overall Integrated Final Risk Judgement can be derived by combining the

rating of the SAPROF with the results of the HCR-20 or another risk measure using structured clinical judgement. Ideally, this should be done by the same rater.

Studies reveal that the SAPROF has good internal reliability (Cronbach's α= 0.88; de Vries Robbe et al., 2011) and has been shown to have good predictive validity for desistance from violent offending after a 1, 2 and 3-year follow-up (AUC = .85, AUC = .80, AUC = .74; de Vries Robbe et al., 2011).

START. The Short-term Assessment of Risk and Treatability (START; Webster et al., 2004) is a tool designed to examine multiple risk outcomes (risk to others, self-harm, suicide, unauthorized leave, self-neglect, substance abuse and victimization by others) and protective factors. A total of 20 dynamic items are coded as both strengths and vulnerabilities, each rated on a 3-point scale. The maximum total score on the START is 80, 40 on each scale. Items on the START include: social skills, relationships, occupational, recreational, self-care, mental state, emotional state, substance use, impulse control, external triggers, social support, material resources, attitudes, medication adherence, rule adherence, conduct, insight, plans, coping, and treatability.

The majority of the items on the START were scored based on overlapping items on the SAPROF. Some items were scored based on the patient's file information since they were not included on the SAPROF. These included mental state (presence of psychotic symptoms), emotional state (current mood), rule adherence (ability to obey rules), and conduct (cooperative with others). Other items were scored on the basis of interview information because they were unlikely to be included in file information, including relationships (e.g., therapeutic alliances with staff and patients) and external triggers (e.g., appropriate current living circumstances).

Since the START is an SPJ tool, patients are coded as low, moderate or high. The START is future oriented and intended for short-term predictions of risks and strengths. Also, the START is highly dependent on interdisciplinary collaboration for assessment and treatment planning by clinicians (Nicholls, Brink, Desmarais, Webster, & Martin, 2006). A recent study demonstrated good predicative validity for the risk scale of the START for interpersonal violence (AUC =.63 to .79), verbal threats (AUC = .66 to .83), and any patient aggression (AUC = .65 to .83) for the short-term, ranging from 30 to 90 days for risk scales only (Chu et al., 2011).

HCR-20. Historical, Clinical, Risk Management (HCR-20; Webster, Douglas, Eaves, & Hart, 1997) is another SPJ tool used to predict violence risk. The most recent version of the measure, HCR-20 Version 3 (HCR-20 V3: Douglas, Hart, Webster, & Belfrage, 2013), was used for the current study. It contains 20 risk factors, 10 are historical items, 5 are dynamic clinical factors and 5 are dynamic risk management factors. Historical items on the HCR-20 V3 are past problems with: previous violence, other antisocial behaviour, relationships, employment, substance use, major mental disorder, personality disorder, traumatic experiences, violent attitudes, and treatment or supervision response. Clinical items on the HCR-20 V3 include recent problems with: insight, violent ideation or intent, symptoms of major mental disorder, instability, and treatment or supervision response. Risk management items on the HCR-20 V3 include future problems with: professional services and plans, living situation, personal support, treatment or supervision response, and stress or coping.

Items are scored on a 3-point scale (0, 1, or 2), and some items are further divided into sub-items. A score of 2 is indicative of clear evidence that the risk factor is present, 1 being sufficient evidence that the risk factor is partially present, and 0 indicative of no evidence that the risk factor is present. Every participant had an updated HCR-20 V3 scoring template

completed by a clinical forensic psychologist following their risk assessment interview. Items are therefore scored based on responses from a semi-structured interview as well as file information.

On the HCR-20 V3, clinicians score each item based on their presence and relevance. Only the presence of items were used for the current study; the relevance of items were disregarded. Scoring templates were collected from clinical forensic psychologists and individual risk item scores were recorded. H scores, C scores, and R scores were computed by summing items under each category. Of note, in clinical practice, item scores are not tallied but rather an overall judgment is assigned given that the HCR-20 is an SPJ tool. Each participant's hospital report was reviewed to record their overall risk rating on the HCR-20 V3 for the upcoming year, should the patient be transferred to a lower security level.

The HCR-20 is a widely used tool for civilly committed patients. It has been shown to have good interrater reliability (ICC=.80) and good predictive validity for criminal violence (AUC=.80) (Douglas, Ogloff, Nicholls, & Grant, 1999). The HCR-20 has been shown to have moderate to large associations for predicting past violence, treatment progress and reduced violence (Douglas, Blanchard, Guy, Reeves, & Weir, 2010; Douglas & Webster, 1999).

LS/CMI. The Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2006) is a risk tool that measures the risks and needs of late adolescence and adult offenders, as well as providing case management. The tool is aligned with the RNR framework. The LS/CMI consists of 43 items with 10 comprehensive sections to assist in offender management. Sections of the measure relate to the primary criminogenic needs: criminal history, education/employment, family/marital, leisure/recreation, companions, alcohol/drug problem, procriminal attitude/orientation and antisocial pattern. A section identifying strengths is also

available on the measure. Items are scored based on responses from a semi-structured interview and file review. Items on the LS/CMI are tallied to produce a final score, which relates to a risk category of low, moderate and high risk of general recidivism. Further, total scores are normed and correspond with a probability of reoffending over a given period of time.

Scores on the LS/CMI were scored by clinical forensic psychologists based on information from the risk assessment interview and file information. Of note, the LS/CMI is only scored on patients who have committed prior offences. Items marked as strengths on the measure were recorded, as well as the overall LS/CMI score and corresponding risk category. Further, scores were recorded on the following 8 scales: Criminal History, Education/Employment, Family/Marital, Leisure/Recreation, Companions, Alcohol/Drug Problems, Antisocial Patterns, Procriminal Attitude Orientation.

Earlier versions of the LS/CMI (Level of Service Inventory-Revised; LSI-R) have demonstrated high internal consistency on the general risk/need items (Cronbach's α=.91) (Girard & Wormith, 2004). Among correctional samples, the LSI-R has been shown to be a good predictor of general offending (AUC=.73) and violent reoffending (AUC=.68) (Girard & Wormith, 2004).

PCL-R. The Psychopathy Checklist-Revised (PCL-R; Hare, 2003) assesses the personality trait psychopathy, although it is also widely used as a risk tool to predict violent and sexual recidivism (Hanson & Morton-Bourgon, 2004; Hare, Clark, Grann, & Thornton, 2000; Leistico, Salekin, DeCoster, & Rogers, 2008). The PCL-R is scored based on a semi-structured interview and file review. It comprises 20 items rated on a 3-point scale (0, 1, or 2). Items on the PCL-R are organized into two factors; Factor 1 represents callous, unemotional, and remorseless features of psychopathy, tapping into the individual's personality structure. Factor 2 represents

the impulsive, antisocial lifestyle characteristics, based on the behavioural aspects and lifestyle characteristics of psychopathy (Harpur, Hare, & Hakstian, 1989).

Total scores on the PCL-R are out of 40. A score of 30 or greater signifies the presence of psychopathy (Hare, 1991); however, others have suggested that scores in the range of 25 to 29 indicate threshold psychopathy (Quinsey, Harris, Rice, & Cormier, 1998). The PCL-R was scored by clinical forensic psychologists for each participant based on the risk assessment interview and file information. PCL-R scoring sheets were collected and scores on both Factor 1 and Factor 2 were recorded for the study.

The internal consistency of the PCL-R was found to be good, ranging from Cronbach's α=.85 to .89 (Hare et al., 1990). Further, the PCL-R was shown to be associated with sexual recidivism (AUC=.70) and violent recidivism (AUC=.75) for forensic psychiatric populations (Barbaree, Seto, Langton, & Peacock, 2001; Hare et al., 2000).

WASI-II. Wechsler Abbreviated Scale of Intelligence (WASI–II; Wechsler, 2011) is an abbreviated version of the WAIS and provides an estimate of IQ. A composite score is calculated using the Verbal Comprehension Index (VCI) and Perceptual Reasoning Index (PRI) as estimates of IQ. The WASI-II comprises 4 subtests: Block Design, Vocabulary, Matrix Reasoning and Similarities. Scores on the WASI subtests are tabulated and normed to compute Full Scale IQ (FSIQ) scores. FSIQ scores of 100 denote average intellectual abilities. After the administration of the WASI-II, the VCI, PRI, FSIQ and percentile ranks were computed for each participant. The WASI has been shown to have good convergent and divergent validity and is therefore a valid screening measure of verbal performance and general intellectual abilities (Hays, Reas, & Shaw, 2002).

SANS/SAPS. Scale for the Assessment of Negative Symptoms (SANS) and Scale for the Assessment of Positive Symptoms (SAPS) (Andreasen, 1983; Andreasen, 1984), is a tool used to assess negative and positive symptoms of psychopathology, specifically symptoms of schizophrenia (Corcoran & Fischer, 1987). Negative symptoms of schizophrenia are often described as deficits in behaviour, whereas positive symptoms of schizophrenia are often described as excesses of behaviour or sensory experiences (Schuldberg, Quinlan, Morgenstern & Glazer, 1990). The SANS and SAPS are rating scales that can be scored using a semi-structured interview to asses both verbal responses and to observe nonverbal cues. The SANS/SAPS can also be completed based on direct clinical observations, observations by family members or reports from treatment team members (Corcoran & Fischer, 1987).

The SANS consists of 25 items, which are divided into five subscales: affective flattening (impoverished emotional expression, reaction and feeling), alogia (impoverished thinking and cognition), avolition and apathy (lack of energy, drive and interest), anhedonia and asociality (lack of interest or pleasure), and impairment of attention (difficulty focusing). The SAPS consists of 35 items, which are divided into four subscales: hallucinations (abnormality in perception), delusions (abnormality in content of thought), bizarreness (behaviour is unusual, bizarre or fantastic), and positive thought formation (fluent speech that tends to communicate poorly for a variety of reasons like distraction). Each item is rated on a 5-point scale: Not at all, questionable, mild, moderate, marked or severe. Symptoms are rated within the last month. While most items on the SANS/SAPS were scored during the research interview, some items were scored immediately following the interview because they relied on the observation of nonverbal cues and patterns of speech (e.g., assessing whether the individual displayed unchanging facial expressions).

The SANS/SAPS Composite score is out of a maximum of 300; scores for the SANS range from 0-125; scores for the SAPS range from 0-175. Higher scores are indicative of greater positive and/or negative symptoms (Corcoran & Fischer, 1987). The SANS/SAPS can be scored in a variety of ways (Corcoran & Fischer, 1987). For the purpose of the current study, total scores on the SAPS and the SANS were computed by adding scores on the global rating scales for each measure. Composite SAPS/SANS were computed by adding all nine global rating scales scores.

Previous administrations of the SAPS/SANS on a large sample of outpatients diagnosed with schizophrenia had a mean score of 32.4 on the SANS (*SD*=15.9) and a mean score of 18.9 on the SAPS (*SD*=12.9) (Schuldberg et al., 1990). The internal consistency was moderate for both the SANS (Cronhbach's α =0.47) and the SAPS (Cronhbach's α =0.58). Inter-rater reliability for the SANS/ SAPS ranged from good to excellent (ICC= 0.83- 0.92) (Schuldberg et al., 1990).

GAF. Global Assessment of Functioning (GAF; American Psychiatric Association, 2000) is a scale within the DSM-IV-TR that ranges from 0 to 100. It is scored by mental health clinicians to provide a rating of social, occupational and psychological functioning. Although the scale is no longer included in the DSM-5, the GAF remains widely used by clinicians (Aas, 2010). The GAF was rated by the researcher for every participant based on the individual's social and occupational functioning at the time of the research interview.

Study Variables

Two outcome variables were examined in the current study: Proxies of recidivism and risk management decisions.

Proxies of Recidivism. Proxies of recidivism comprise three variables: PRN medication, institutional misconduct, and disposition breaches over the previous year and over a six-month follow-up. The rationale for using a six-month follow-up period is that six-months is considered a "medium-term" prediction and is the minimum prediction period outlined for the SAPROF (Coid, Kallis, Doyle, Shaw, & Ullrich, 2015; de Vries Robbe, 2014). Further, a six-month follow-up was shown to be a sufficient period of time to predict institutional behaviour using the SAVRY, another protective factor instrument (Gammelgard, Koivisto, Eronen, & Kaltiala-Heino, 2008). Not all participants were included in the six-month follow-up. One reason for the varying sample size at follow-up was that some participants received an absolute discharge and were no longer under the authority of the ORB and their information was not available. Another reason for the varying sample size was that some participants were transferred to higher levels of security, either a prison/maximum security facility, and so their information could not be accessed.

Information about PRN administration was logged in the patient's file and psychiatrists' notes on file. Only PRN administration for psychiatric medications was included: antidepressants, anxiolytics or antipsychotics. The reason for the PRN administration was reviewed and considered, which was due to either behavioural management issues or symptom complaints by the patient. Specifically, reasons included anxiety, agitation, aggressiveness, violence towards self and/or others. This variable was coded as either absent or present, and the number of PRN administrations was tallied.

The second proxy of recidivism was institutional misconduct among inpatients only. Institutional misconduct was broadly defined as inappropriate behaviour or behaviour in opposition to hospital rules. This variable was further divided into two types; major and minor

misconducts. Minor institutional misconducts involved returning late from grounds privileges greater than five minutes. Major institutional misconducts involved smuggling contraband, smoking cigarettes on the unit, failure to follow rules, aggressive behaviour towards patients or staff, being verbally threatening and physical violence towards staff or patients. The presence and number of major and minor misconducts were calculated. The result of institutional misconduct may involve escorting the patient to a seclusion room for a period of time and/or criminal charges. Information regarding institutional misconduct was logged in patients' files.

Disposition breaches were the third proxy of recidivism, defined as a violation in the conditions outlined by the ORB. Disposition breaches were categorized for both inpatients and outpatients as: Eloping from the hospital, using illicit substances, physical aggression towards others, possessing weapons and refusing psychiatric medication. Of note, since conditions outlined in patients' dispositions vary, what is considered to be a disposition breach may vary across patients. Disposition breaches committed by outpatients were usually brought to the attention of staff and clinicians monitoring the patient, who then report the occurrence to the FOS at Ontario Shores. The police are then contacted for situations involving violence, which may result in scheduling an early ORB hearing. The presence and number of disposition breaches were obtained from information in incident reports and file notes.

Overall, information about proxies of recidivism were based on a review of participants' file. To control for varying lengths of institutionalization, only information from the past year was included. Proxies of recidivism were also measured six-months following the participants' interview, based on file information.

Risk management decisions. Risk management decisions consist of two factors: The level of condition/privilege and level of security granted. There are 14 possible privilege levels, refer

to Appendix B for a list of ORB privilege/condition levels. Both inpatients and outpatients were included for this analysis, since outpatients can receive a change in their level of privilege. Specifically, outpatients could receive fewer privileges if they were to be transferred to a higher level of security and return to their inpatient status. This variable was coded in three ways: an increase in security or privilege level (change scores), a decrease in security or privilege level (change scores) or no change.

Higher privilege levels indicate less conditions and greater personal liberties. Patients residing on medium and minimum security units are entitled to different privileges. Patients on medium units can only receive privileges for medical, escorted, accompanied, or with an approved person. Therefore, these privileges correspond with either level 1 or 8, whereas patients residing on minimum units can receive any level ranging from 1-14. Levels 1-7 concern hospital and ground conditions. Level 1 involves the patient being directly accompanied for mandatory medical and legal circumstances, as well as compassionate uses. Levels 8-14 concern community conditions. Level 14 involves living in the community on a detention order. Patients with a conditional discharge who are living in the community are not assigned a privilege level.

The maximum privilege level is stipulated by the ORB and outlined in the patient's disposition. The treatment team then determines what level should be assigned to the patient without exceeding this maximum level. Therefore, this process can be somewhat unstandardized and at the discretion of the treatment team. Throughout the year, patients submit a request form to their treatment team to apply to increase their privilege level. There are no limits to the number of times a request form can be submitted. The treatment team usually encourages individuals to apply for privileges, especially for those patients who do not advocate for themselves. However, some patients may not take advantage of their assigned privilege level.

Further, privilege levels may not be assigned in sequential order, rather assigned based on what is needed by the patient (e.g., the patient only requires 2 hours of indirect supervision in the hospital so they may receive a level 4, although they are permitted to receive a higher level). Privilege change was coded based on changes outlined in the patient's new disposition compared to their previous disposition (the highest privilege outlined in their disposition). Ultimately, privilege level depends on the ORB, but the treatment team can dictate the exact the level based on a range outlined by the ORB.

Participants were also assessed on the level of security granted in their ORB disposition (e.g., medium security, hybrid order, minimum security without community living, minimum security with community living, conditional discharge, absolute discharge). Hybrid orders are assigned by the ORB and give the treatment team the discretion to determine the best inpatient setting for the patient. Therefore, the patient's disposition will state both medium and minimum levels of security.

Changes between levels of security granted by the ORB were assessed by examining whether a change in the level of security was present and the direction of change before and after the ORB hearing. This was conducted by rank ordering disposition types by the level of freedom permitted. Each level of security was assigned a number from 1-7, higher numbers are associated with greater security or supervision and less freedoms. Change scores were assigned to determine an increase, decrease or no change in security level. Positive change scores denote a transfer to a lower level of security (e.g., minimum to conditional discharge). Negative change scores denote a transfer to a higher level of security (e.g., minimum to medium). Participants who were not granted a change or stayed at the same level of security received a score of 0 (see Table 1).

Of note, the difference between supervision levels is not equal. For instance, there are more minimum disposition types than maximum disposition types. As such, the magnitude of change between security levels could not be evaluated to reflect differences in the level of security permitted in a clinical setting.

Table 1

Direction of Change in Level of Security

Disposition before ORB	Disposition after ORB							
	Maximum security	Medium security	Hybrid disposition	Minimum security-NOCL	Minimum security - CLIH	Minimum security - CLIC	Conditional discharge	Absolute discharge
Maximum security	0	1	2	3	4	5	6	7
Medium security	-1	0	1	2	3	4	5	6
Hybrid disposition	-2	-1	0	1	2	3	4	5
Minimum security-NOCL	-3	-2	-1	0	1	2	3	4
Minimum security - CLIH	-4	-3	-2	-1	0	1	2	3
Minimum security - CLIC	-5	-4	-3	-2	-1	0	1	2
Conditional discharge	-6	-5	-4	-3	-2	-1	0	1
Absolute discharge	-7	-6	-5	-4	-3	-2	-1	0

Note. NOCL= No community living. CLIH=Community living but residing in the hospital. CLIC=Community living and residing in the community. 0=no change

Predictors. The primary predictor variables under investigation were: 1) SAPROF total scores, and Internal, Motivational, and External subscale scores; 2) HCR-20 total scores; and 3) Index Scores. Total scores on risk and protective factor measures were tallied. Although it is not recommended to calculate total scores for SPJ tools, it is common practice in research (de Vries Robbe, 2014; Douglas & Kropp, 2002). The Index Score was used as a predictor, which was calculated by subtracting the HCR-20 V3 total score from the SAPROF total score, representing corrected risk. The Index Score is a conceptual measure suggested by the developers of the SAPROF to account for the reduction in risk through the presence of protective factors, and considered effective for formulating patient risk (de Vries Robbe, de Vogel, & Douglas, 2013). In essence, the Index Score measures the remaining risk once protection is subtracted.

Procedure

The protocol for the current study was reviewed and approved by the Research Ethics Board at Ryerson University and the Research Ethics Board at Ontario Shores Centre for Mental Health Sciences. A consent form was developed by the researcher and staff at Ontario Shores, detailing the nature, purpose, risks and benefits of participation. Participants were informed that their participation was voluntary and were informed that they could withdraw from the study at any time. Also participants were informed that their participation and responses would not impact their clinical care. Participants were made aware that information provided would be de-identified after the interview to ensure confidentiality, and used for research purposes only. Therefore, information obtained from the study was not released to the ORB.

Participants were made aware that agreeing to participate involved allowing the researcher access to their patient file on Meditech (the hospital's computer database) and access to their past and upcoming ORB dispositions. As described above, special care was taken to

ensure that patients were capable of providing informed consent and thus not actively psychotic at the time of the interview. Patients incapable of providing informed consent were excluded from the study.

Patient recruitment and data collection by the researcher and a research assistant took place from January 2015 to September 2015 (January to October ORBs). Clinical forensic staff and members of the treatment team on each unit (e.g., forensic nurses) were aware of the study. Prior to the implementation of the study, the researcher met with the clinical forensic psychologists to discuss the study and provide them with a recruitment script outlining the study in simple terms.

This study was a prospective cohort study, meaning that participants were evaluated prior to the outcome (i.e., their ORB hearing). Clinical forensic psychologists asked patients if they were willing to be approached by a researcher during the informed consent process of the risk assessment interview. To avoid coercion, clinical forensic psychologists clearly stated to patients that they were not affiliated with the study. Names of interested patients were relayed to the researcher who contacted the patient on their unit and set up an interview at the patient's earliest convenience.

Further, clinical forensic psychologists were advised that if the patient appeared anxious about their upcoming ORB hearing or risk assessment interview, or expressed that they were ambivalent about participating in the study, the patient was not required to provide the psychologist with a definite answer at that time. Instead, patients were encouraged to take some time to consider participation. Afterwards, clinical forensic psychologists were instructed to re-approach the patient at a later date, sometime before their ORB hearing to mention the study again. This procedure was implemented in an effort to increase participation rates, especially

since feedback from clinical forensic psychologists was that recruitment occurred at a particularly stressful time for the patients and that they may be more likely to decline as a result.

The administrator of the Forensic Outpatient Service (FOS), who contacts patients by telephone on a monthly basis to schedule risk assessment interviews, scheduled outpatients at Ontario Shores. The rationale for implementing this protocol was that one of the clinical forensic psychologists designated to the FOS was affiliated with the current study. Therefore, in order to prevent a potential conflict of interest and perception of coercion, the administrative staff was the first point of contact for all outpatients. During the course of the current study, this affiliated psychologist was moved to another unit. Other clinical forensic psychologists recruited patients on her behalf in order to reduce any potential conflict of interest.

Each participant received a $10.00 gift card to Tim Hortons as remuneration in return for 60-90 minutes of participation. A token amount was provided given that coercion may be a factor among patients, many of whom are lower income and considered a vulnerable population.

Once recruited, each patient was individually interviewed by the researcher for approximately 60-90 minutes; some of the interviews were conducted over the course of several sessions for patients who were more symptomatic. For inpatients, the research interview took place in an assessment room on the patient's unit. For outpatients, research interview took place in the outpatient division of the hospital in an interview room. Participant interviews were prioritized on the basis of their scheduled ORB hearings which occur annually. Information regarding scheduled hearings were provided to the researcher by the treatment team to establish which patients are eligible to participate in the study and should be recruited. Participants were only assessed at one-time point, preceding their ORB hearing, thus test-retest reliability of the SAPROF was not evaluated.

After patients were approached and consent was obtained, they were provided the option to conduct the interview at that immediate point in time or schedule the interview later that day or week. A review of participants' files took place after the research interview to ensure that file information did not bias the interview. To ensure that knowledge about risk level did not influence the researcher, ORB dispositions, psychology reports, risk assessment scoring forms and hospital reports were not reviewed until the all measures from the interview battery were scored.

Some outpatients expressed an interest in participating in the study but lived too far to commute to the hospital and did not have monthly routine appointments scheduled at the hospital, as initially anticipated. In order to accommodate these outpatients, they were provided the opportunity to participate in a phone interview, rather than an in-person interview. These patients were recruited the same way as other participants, through the FOS administrator. Informed consent was reviewed over the phone and verbal consent was obtained by the researcher. The participant's FOS clinician (either a nurse or case worker who meets with the patient regularly in the community) had the participant sign the consent form prior to the interview. The FOS clinician was notified upon the completion of the phone interview and provided the participant with the compensation and had the participant sign the proof of compensation receipt the next time they met.

During the informed consent process, participants were asked to consent to voice recording the interview for research purposes only in order to verify the information provided during the interview (e.g., interrater reliability). However, the audio recording was optional and participants' decision not to be recorded did not prevent them from participating in the study. Three participants who agreed to a phone interview did not have their interview audio recorded.

Assessment measures were administered in the following order: SAPROF, START, WASI-II, and SANS/SAPS. The majority of items on the START were scored based on responses from the SAPROF interview guide, due to overlapping items, and required minimal additional questioning. In order to score the first item on the SAPROF, intelligence must be assessed. Although the WASI-II does not include all indexes and subtests of the WAIS-IV, using dyad short forms are appropriate when estimating global intelligence (FSIQ) (Girard, Axelrod, Patel & Crawford, 2015). An interview guide comprising questions used to score the SAPROF was adapted from the SAPROF Interview Self-Appraisal (SAPROF-ISA: de Vries Robbe & de Vogel, 2014). The SAPS/SANS was scored based on a semi-structured interview guide that was developed by a researcher at the Centre for Addiction and Mental Health (CAMH) (Zawadzki, 2014). The GAF was scored following the interview. The researcher was blind to each participant's ORB disposition and proxies of recidivism at the time of scoring the SAPROF. Following the completion of the interview, participants were given the opportunity to ask the researcher any additional questions. The results were mailed to the facility to distribute to participants upon request.

Clinical forensic psychologists administered risk and personality measures, as part of a routine risk assessment prior to the patient's ORB hearing. Specifically, the HCR-20 V3, LS/CMI and PCL-R were scored by clinicians. The overall risk ratings (e.g., low, medium, high) were determined for the HCR-20 V3 by clinicians and individual item scores were tallied for the current study by the researcher based on clinicians' notes and assessment scoring sheets.

Of note, patients who resided at Ontario Shores for less than a year (because they were residing at a different institution or newly found NCRMD), did not have risk assessment

information from the previous year or information about proxies of recidivism. As such, this information was missing for some participants.

Upon completion of the interview and ORB hearings, participants' file information was perused and demographic information was gathered. Additionally, information regarding psychiatric diagnoses, psychiatric medication, institutional misconduct, disposition breaches and risk scores assigned by clinical forensic psychologists prior to their hearing and over the past year were obtained. ORB dispositions, the final decision report constructed by the ORB comprising patient outcomes was also accessed to gather information regarding the security level assigned and conditions within the hospital for the upcoming year. Proxies of recidivism were also gathered six-months following the initial research interview based on file information. The period of time between the risk assessment, research interview and ORB hearings ranged anywhere from several days to a couple of weeks.

All participant information was entered into a computer database and de-identified using a participant number. Hard copies of response forms from the administered measures and consent forms were securely stored in a locked cabinet in an office at Ontario Shores, separate from consent forms. Data will be stored for 7 years, after which it will be shredded and destroyed.

Statistical Analyses

The following analyses were conducted to address research questions stated above.

Inter- and intrarater reliability. To ensure the reliability of SAPROF scores, interrater reliability was established. Interrater reliability was measured by randomly selecting a portion of the administered tests (10-15%) and having another clinician involved in the study at Ontario Shores listen to the audio recordings and score the measure. Therefore, interrater reliability was

only based on a select sample of participants who consented to being voice recorded. Eight participants were selected and scored for interrater reliability. Interrater reliability was based on the information pulled from patient files and information disclosed during the interview, which was all inputted into a SAPROF scoring form. As a result, the second rater used the same information as the first scorer to rescore each item on the SAPROF.

Intrarater reliability of the SAPROF was also assessed to determine whether there was consistency between administrations of the SAPROF by a single rater (the researcher). Eight participants were randomly selected and scored for intrarater reliability. Cronbach's α, a measure of the internal consistency of a scale, was used to assess intra- and interrater reliability for SAPROF total and item scores.

Confirmatory factor analysis. Confirmatory Factor Analysis (CFA) is a multivariate technique that evaluates whether covariations among variables load onto one or more latent variables or factors in the data (Arbuckle, 2009). The reliability of individual items, factors and the overall instrument were assessed by determining whether the current data fit the proposed measurement model (Kline, 2010). A CFA was conducted to confirm the original factor structure of the SAPROF, comprising Internal, Motivational and External subscales. Analysis of Moment Structures (AMOS), Version 18.0 (Arbuckle, 2009) was used to test the fit of the data to the model. Several goodness of fit indices were used to evaluate model fit.

Constraints were imposed in order to establish the scale of the latent variables in the CFA model (Brown, 2015). The factor with the highest loading (the coefficient that was highest when the CFA was first ran) was set to 1.00 for each latent variable. Each factor loading was then adjusted to range from -1.00 and 1.00 relative to the highest factor loading. This criterion was made in order to standardize factor loadings and can be interpreted relative to other factor

loadings. The factor loading signifies the strength of the linear relationship between the latent variable and the error of measurement.

The absolute fit of the model to the data were measured using the ratio of the chi-square to degrees of freedom (χ^2/df). This index represents the difference between observed and expected covariance matrices, and is used to reduce sensitivity to small sample sizes. Values closer to zero are indicative of a better fit, and a χ^2/df ratio less than 2 is indicative of a good fit (Gatignon, 2003). The root mean square error of approximation (RMSEA) is another index of absolute fit that calculates the discrepancy between the hypothesized model and the population covariance matrix (the non-centrality based estimate of error). An RMSEA value less than .06 is representative of a good fit, and < .08 is a reasonable fit (Hooper, Coughlan, & Mullen, 2008; Hu & Bentler, 1999). The associated confidence intervals and PCLOSE value helps to identify sampling error, *p* values greater than .05 (not statistically significant) indicate a close fit between the data and hypothesized model. The comparative fit index (CFI) and normed fit index (NFI) are both relative fit indices that were used. The CFI index examines the discrepancy between the data and hypothesized model by adjusting for the sample size. CFI values greater than .90 or larger are considered representative of acceptable model fit (Hu & Bentler, 1999). NFI index analyzes the discrepancy between the χ^2 of the hypothesized model and χ^2 of the null model, values greater than .95 are indicative of good model fit (Hu & Bentler, 1999).

Correlations. A series of correlations were conducted to establish construct validity of the SAPROF, by investigating both convergent and divergent validity. Pearson and Spearman correlations were conducted between SAPROF total and subscale scores and measures of risk, protection, general functioning and mental state for research questions one and two. Pearson's correlations were used to correlate SAPROF total scores with other continuous variables, while

Spearman's correlations were used to correlate SAPROF judgement ratings with other categorical measures.

For correlation coefficients, *r*=.1 signifies a small effect size (explaining 1% of the total variance), *r* = .3 signifies a medium effect size (accounting for 9% of the total variance), and *r* = .5 signifies a large effect size (explaining 25% of the total variance) (Cohen, 1992). In order to reduce Type 1 error, only correlations with at least a large effect were considered for convergent validity between two measures. Further, only correlations less than a medium effect were considered to determine divergent validity between two measures. There was no mention of established thresholds for determining convergent and divergent validity in previous SAPROF validation studies (Abidin et al., 2013).

In addition to calculating correlations between measures, a correction for attenuation was applied for total scores only, since the reliability of the full measure is required for this calculation. Correction for attenuation or disattenuation is a statistical procedure used to account for the weakening effect of measurement error (Spearman, 1904). Correction for attenuation is calculated using the following formula: r*xy* /sqrt (r*xx* x r*yy*), where *xy* represents the correlation between the SAPROF and another measure, *x* represents the reliability of the SAPROF and *y* represents the reliability of the other measure. Values greater than 1.00 indicate that measurement error is not randomly distributed and suggests that two tests are measuring the same trait (Muchinsky, 1996).

Regression analyses. In accordance with research question three, multiple regression analyses were conducted to determine whether SAPROF subscale scores predicted the absence/presence of PRN administrations, institutional misconduct and disposition breaches. Also, logistic regression analyses were conducted to determine whether SAPROF subscale

scores predicted an absence/presence of PRN administrations, institutional misconduct and disposition breaches.

Both logistic and multinomial regressions were employed to address research question four. Logistic regression analyses were conducted to determine whether SAPROF total scores predicted a change in privilege level and a change in security level. Next, multinomial regression analyses were conducted to examine whether SAPROF total scores predicted the direction of change (increase in privilege/security, decrease in privilege/security or no change).

Hierarchical multinomial logistic regression analyses were employed to identify how protective factors operate in relation to risk factors, as proposed in research question five. HCR-20 V3 total scores were entered as the first step and SAPROF total scores were entered in the second step to determine incremental predictive validity. Outcome variables were dichotomous, and included the presence of PRN administrations, institutional misconduct, disposition breaches, increase in privilege level and decrease in security level.

Given the number of patient characteristics collected and examined for research question six, stepwise logistic regression analyses were conducted. Stepwise regression analyses are often used in the exploratory stages of model building to determine which predictors are most useful (Tabachnik & Fidell, 2000; Thayer, 2002). A conditional forward method was used. This statistical process involves systematically adding the most significant variables and removing the least significant variable during each step. The probability for inclusion was .05 and .10 for the exclusion of variables. All predictors and interaction terms were entered together to control for alpha levels and prevent inflation.

In addition to SAPROF total scores, the following predictors were entered into the stepwise logistic regression: the type of patient (inpatient vs. outpatient), length of patient

institutionalization (in years), length of time on current disposition (in years), severity of the index offence (severe/contact offence or not), number of psychological treatment sessions attended (individual and group) and participation in vocational activities (volunteer inside and/or outside of the hospital. The selection of these variables were clinically derived and therefore, may potentially influence risk management decisions and proxies of recidivism.

A median-split was calculated to categorize patients as either institutionalized for more or less than 3.5 years. Interaction terms were entered by multiplying each patient characteristic with SAPROF total scores. Continuous predictors were centred (subtracting each score from the mean) before creating the interaction term to reduce the possibility of multicollinearity (Robinson & Shumacker, 2009). Five dichotomous outcome variables were predicted for the stepwise logistic regression analyses: PRN administrations, institutional misconduct, disposition breaches, increase in privilege level and decrease in security level.

Receiver Operating Characteristic (ROC) Analysis. ROC analysis is an analytical technique that produces an Area under the Curve (AUC) value. The ROC curve is created by plotting the true positive rate (hit rate) against the false positive rate (probability of false alarm rate) (Swets 1988). This method is independent of the number of events/incidents of the outcome (e.g., base rates). An AUC value of 0.5 represents a chance level prediction, whereas 1.0 perfectly predicts the outcome variable. AUC values of .71 are considered to be equivalent to Cohen's d of .80, indicative of a very large statistical effect size (Douglas, Yeomans, & Boer, 2005; Dunlap, 1999; Rice & Harris, 2005). ROC analyses were conducted to determine AUC values of the HCR-20 V3 total score, SAPROF total score, and Index Score across five dichotomous outcome variables: PRN administrations, institutional misconduct, disposition breaches, increase in privilege level and decrease in security level.

Pairwise comparisons were employed between the three test variables to determine whether AUC values were statistically significant. Pairwise comparisons were carried out using the MedCalc software. The DeLong, DeLong & Clarke-Pearson (1988) approach was used for the calculation of the standard error of the AUC value and differences between the two AUCs. This is the recommended approach and does not make parametric assumptions of the data (Cleves, 2002). The calculation applied was: $z = A_1\text{-}A_2 / \text{sqrt } V(A_1\text{-}A_2)$, where A represents the AUC value and V signifies the variance. The denominator of the equation is calculated as $V(A_1) + V(A_2) - 2\,\text{Cov}(A_1, A_2)$ (DeLong et al., 1988).

Statistical Assumptions

Statistical assumptions were tested for multiple linear regression and logistic regression (Field, 2009). For multiple linear regression, linearity and homoscedasticity assumptions (error terms along the regression are equal) were assessed. A scatterplot was produced in order to detect randomness of data and ensure that no outliers were present. Two additional graphical methods were used to provide an indication of whether the residuals formed a normal distribution. First, a histogram of the residuals was examined and it was determined that the histogram and curve were similar. Second, a q-q plot was examined and it was determined that the points fell close to the diagonal line. As such, linearity and homoscedasticity were not violated.

Next, multicollinearity was assessed using various methods, including the tolerance and variance inflation factor (VIF) statistic. Tolerance values were > 0.2, signifying that multicollinearity was not present in the data. The average VIF value was 1.6, suggesting that multicollinearity may be biasing the regression model, however VIF values >10 is considered a real cause for concern (Bowerman & O'Connell, 1990; Myers, 1990). To further assess

multicollinearity, correlations between independent variables were conducted using Pearson's correlation. Although both Internal and Motivational subscales on the SAPROF were positively correlated (r =.58, p <.01), correlation coefficients between predictors were less than .80. Therefore, multicollinearity was not a concern and no further adjustments were necessary.

Assumptions were also tested for logistic regression (Field, 2009). Although logistic regression does not assume linearity, it does assume a linear relationship between continuous predictors and the logit of the outcome variable. This assumption was tested by examining whether the interaction term between the predictor and its log transformation is significant. As outlined by Hosmer, Lemeshow, and Sturdivant (2000), the interaction between continuous predictors and their natural logarithms should not be significant. Interaction terms were created for HCR-20 V3 total scores and SAPROF total scores. None of the interaction terms were significant and the assumption of linearity was met (Field, 2009).

All correlation coefficients were below .80, which indicates that multicollinearity (e.g., signifying that two or more variables are highly correlated) was not present. Further, the tolerance and VIF statistics were assessed, tolerance values were > 0.2 and VIF values were < 10 indicating the absence of multicollinearity in the data and that the relationship between variables were not problematic. Lastly, the independence of errors, which specifies that cases are independent and that the same participants are not measured at different points in time, was confirmed. Given that each participant was assessed at one-time point, the independence of errors assumption was satisfied.

Alpha Criterion

Given the relatively small sample size and the number of analyses conducted, an alpha of .01 was applied a priori. The increased number of statistical analyses may lead to a greater

chance of Type 1 error, obtaining a false-positive result. Rather than using the conventional .05 alpha, a more conservative alpha criterion was employed. A full Bonferroni adjustment using family-wise comparisons was not used for concern that the criterion was too stringent and that the presence of an effect would not be detected (Kazdin, 2003). The current study may be considered exploratory in nature, requiring several comparisons to be performed for each hypothesis. Further, the replication of findings is expected with exploratory research, which could safeguard against the effects of a false-positive result. Therefore, in order to balance the risk of Type 1 and Type 2 error rates, the current study used a $p < .01$ cut-off for all tests. Additionally, effects sizes were reported to determine the magnitude or strength of the relationships/predictions (e.g. correlation coefficients and Nagelkerke's pseudo R^2) (Kazdin, 2003).

Results

Sample Descriptives

Of approximately 250 forensic patients at Ontario Shores, 188 were eligible and approached to participate in the study, as their ORB hearing fell within the data collection period January to October 2015. Sixty-two patients had their ORB hearing in November and December which took place after data collection and were not approached to participate. Fifty of the 188 eligible patients (27%) agreed to participate and completed the study. The remaining 138 (73%) either declined to participate or were excluded because they were incapable of providing informed consent. Of the 50 participants who agreed to participate, only one withdrew early from the study, but provided written consent to use their data in the study. Participants' average age was 41.13 years (*SD*=12.57) ranging from 21 to 70 years. The majority of participants were males (*n*= 43, 86%). Participant characteristics are presented in Table 2.

All participants were assigned a psychiatric diagnosis and most participants had multiple diagnoses (see Table 2). The nature of the index offence was scored dichotomously based on whether the offence was violent or non-violent. The Cormier-Lang Criminal History Scores for Violent Offenses and The Cormier-Lang Criminal History Scores for Nonviolent Offenses Scales (Quinsey, Harris, Rice & Cormier, 2006) were used to differentiate the severity of index offence types. More severe contact offences involving violence were coded as violent offences and the remainder were coded as nonviolent offences. The majority of participants (*n* = 34, 68%) committed a violent index offence. For cases in which participants had multiple charges for their index offence, only the most severe violent offence was used for purposes of this study.

Participants' current work status was scored using the *work* item on the SAPROF. From the total sample, 27 (54%) participants were employed, volunteered, or enrolled in school on

either a part-time or full-time basis over the past six-months to a year. Similarly, participants' current intimate relationship status was scored using the *intimate relationship* item on the SAPROF. One (2%) participant was either in a relationship or in a relationship that was stable or supportive over the last six-months to one year.

Participants varied in the length of time hospitalized at Ontario Shores (M = 5.11 years, SD = 4.33), spanning from 0 years to 17.34 years. The mean length of time between being found NCRMD and the interview for this study was 7.31 years (SD = 8.20). The length of time participants were placed on their current disposition ranged from 3 months to 10 years (M = 2.67 years, SD = 2.36).

Table 2

Frequency of Participant Characteristics (N = 50)

Characteristic	*n*	Percent
Fully or partially employed/volunteering/enrolled in school	27	54%
Stable/serious intimate relationship	1	2%
Diagnoses[1]		
Schizophrenia/Psychotic Spectrum	46	92%
Substance Dependence	32	64%
Personality Disorder/traits	22	44%
Mood Disorder	12	24%
Developmental Disorder	5	10%
Anxiety Disorder	3	6%
Impulse Control Disorder	2	4%
Sexual Disorder	1	2%
Prior Offences as Adult	44	88%
Juvenile Record	20	40%
Nature of Index Offence		

Assault/ aggravated assault	16	32%
Murder	8	16%
Attempted murder	4	8%
Uttering threat	7	14%
Attempted or committed robbery	6	12%
Arson	3	6%
Fail to comply with probation	3	6%
Harassment/stalking	2	4%
Mischief	1	2%
Disposition Types (year 2014-2015)		
Secure (medium security)	15	30%
Hybrid Order	5	10%
General Forensic without community living in the hospital (minimum security)	3	6%
General Forensic with community living in the hospital (minimum security)	6	12%
General Forensic with community living in the community (minimum security)	11	22%
Conditional discharge	10	20%
Patient Units at Ontario Shores		
Forensic Assessment Unit (medium)	3	6%
Forensic Assessment and Rehabilitation Unit (medium)	7	14%
Forensic Rehabilitation Unit (medium)	9	18%
Forensic Community Reintegration Unit (minimum)	7	14%
Forensic Psychiatric Rehabilitation Unit (minimum)	4	8%
Forensic Transitional Unit (minimum)	2	4%
Forensic Outpatient Service	18	36%

[1]*Note.* Due to comorbidities, this category does not add up to 100%. For the majority of participants, Schizophrenia/Psychotic Spectrum Disorders were the primary diagnosis.

Total scores and judgement ratings. Three different scores were generated from the SAPROF: SAPROF total scores (summing all item scores of the measure), subscale scores

(summing items by subscale) and judgement ratings (assigning a category based on overall level of protection), based on information collected over the past year (see Table 3). SAPROF total scores ranged from 9 to 32, out of 34. The distribution of SAPROF total scores was assessed and found to be within normal limits (skewness = .20, kurtosis = -.09). SAPROF total scores were higher for inpatients (*M* =19.69, *SD* = 5.46) than outpatients (*M* = 17.22, *SD* = 4.28); however, this difference was not statistically significant, *t* (48) = -1.77, *p* = .09. SAPROF subscale scores were also found to be within normal limits: Internal subscale (skewness = .29, kurtosis = -.80), Motivational subscale (skewness = -.23, kurtosis = -.05), and External subscale (skewness = -.37, kurtosis = -1.13).

SAPROF judgement ratings ranged from low (*n*= 4, 8%), low-moderate (*n*=15, 30%), moderate (*n* =15, 30%), moderate-high (*n* = 10, 20%), and high (*n* = 6, 12%). The distribution of SAPROF judgement ratings were further assessed and also within normal limits (skewness = 0.17, kurtosis= -0.71). The modal category for SAPROF judgement ratings was a moderate level of protection.

Table 3

Mean SAPROF Scores

SAPROF Scale	*M*	*SD*
Total score	18.80	5.16
Internal subscale	4.86	2.46
Motivational subscale	8.36	2.83
External subscale	5.58	2.06

Note. N=50. SAPROF total score is out of 34, Internal subscale score is out of 10, Motivational subscale score is out of 14, and External subscale score is out of 10.

Other measures were also administered during the current study, including the START, SANS/SAPS, WASI-II, and GAF (refer to Table 4). The START judgement ratings ranged from low (*n* =10, 20%), low-moderate (*n* = 17, 34%), moderate (*n* = 16, 32%), moderate-high (*n* = 6,

12%), high (*n* = 1, 2%). The distribution of scores for the START Risk Scale (skewness = 0.30, kurtosis = -0.52), and START Strength Scale (skewness = -0.35, kurtosis = -0.49) was within normal limits. The distribution of scores for the Composite SANS/SAPS score (skewness = 1.07, kurtosis = 0.71) and GAF (skewness = -.23, kurtosis = -.78) was also within normal limits. The modal score for the FSIQ on the WASI-II was 80.

Table 4

Means of Other Measures and Scales Administered

Measure	*M*	*SD*
START		
Total Score	39.66	1.77
Risk Scale Total	15.26	8.23
Strength Scale Total	24.40	8.41
SANS/SAPS		
SANS Score	27.70	19.44
SAPS Score	17.70	19.56
Composite Score	43.56	30.75
GAF	50	14.83
FSIQ	93.67	22.51

Note. *N*=50 for all scores except the FSIQ which was *N*=30. The START total score is out of 80, the Risk scale score is out of 40 and the Strength scale score is out of 40. SANS score is out of 125, SAPS score is out of 175, and the Composite SANS/SAPS score is out of 300. GAF score is out of 100. FSIQ average score ranges from 85-115.

Preliminary Analyses

Risk measures. The HCR-20 V3, LS/CMI, and PCL-R were scored by clinical forensic psychologists as part of the patient's routine risk assessments (see Table 5). The average HCR-20 V3 total score was 21.63 (*SD* = 7.21) out of a possible score of 40. Participants were also assigned a risk rating by the clinical forensic psychologist, based on their predicted level of risk should the patient move to a lower level of security within the upcoming year. The range of HCR-20 V3 risk ratings was low (*n* = 4, 8%), low-moderate (*n* = 15, 30%), moderate (*n* = 15,

30%), high-moderate (*n* = 10, 20%), and high (*n* = 6, 12%). The distribution of scores for the HCR-20 V3 were within normal limits (skewness = 0.01, kurtosis = -0.92). The average Index Score was 3.3 (*SD* = 10.85), indicative of an excess of risk factors. The distribution of LS/CMI categories was very low (*n* = 1, 2%), low (*n* = 7, 14%), medium (*n* = 19, 38%), high (*n* = 12, 24%). The distribution of scores for the LS/CMI (skewness = 0.05, kurtosis = -.69) and the PCL-R (skewness = 0.55, kurtosis = -0.29) were both within normal limits.

Table 5

Mean Scores for Risk Measures

Measure	*N*	*M*	*SD*
HCR-20 V3			
Total Score	50	22.10	7.24
H items	50	13.66	3.46
C items	50	4.14	2.89
R items	50	4.34	2.91
LS/CMI			
Total Score	39	16.36	6.29
Criminal History (CH)	39	2.26	1.21
Education/employment (EE)	39	1.41	1.12
Family/marital (FM)	38	1.11	0.98
Leisure/recreation (LR)	39	1.64	1.37
Companions (CO)	38	1.79	1.17
Alcohol/drugs problems (ADP)	39	1.15	1.04
Procriminal attitude/orientation (PA)	39	1.03	0.96
Antisocial pattern (AP)	39	1.51	1.12
PCL-R			
Total Score	47	13.42	6.59
Factor 1	40	3.90	3.52
Factor 2	40	7.48	3.42

Note. The HCR-20 V3 total score is out of 40, Historical items are out of 20, Clinical items are out of 10 and Risk Management items are out of 10. The LS/CMI total score is out of 43, CH scale is out of 8, EE scale is out of 9, FM scale is out of 4, LR scale is out of 2, CO scale is out of 4, ADP scale is out of 8, PA scale is out of 4, and AP scale is out of 4. The PCL-R is out of 40, Factor 1 has 8 items and Factor 2 has 9 items.

Reliability Analyses

Intrarater reliability. The degree of agreement between two scorings of the SAPROF (Time 1 and Time 2) by one rater was calculated in October 2015. A random subsample (n = 13, 26%) of the study sample was selected using a random number generator. The rate of agreement was assessed between Time 1 and Time 2 for the 17 SAPROF items, and the SAPROF total score (see Table 6). The median time between the first and second rating of the SAPROF was approximately 6 months. The agreement ranged from fair (Kappa = .21-.40) to nearly perfect/perfect (Kappa = .81-1.00), in accordance with Landis and Koch's (1977) thresholds. The average Kappa value (M = .55, SD = .17) was indicative of moderate agreement within the rater over two time periods, which is within the accepted threshold (Landis & Koch, 1977). Item 13 (*social support*) had the lowest intrarater reliability and item 17 (*external control*) had the highest intrarater reliability. Almost perfect agreement between scoring periods was found for SAPROF total scores (r =.95, p <.001).

Interrater reliability. Interrater reliability was calculated upon completion of the data collection, in November 2015. Eight (16%) participants were randomly selected, using a random number generator. The rate of agreement between two raters was assessed for SAPROF scores coded at Time 1 by the researcher and scores coded by a second rater. There was a wide range in agreement between scores, spanning from slight (Kappa = .00-.20) to almost perfect (Kappa = .81-1.00), according to established thresholds (Landis & Koch, 1977). The average Kappa (M = .30, SD = .23) was indicative of fair agreement between raters. Comparable to intrarater scores, item 17 (*external control*) had the highest reliability. Rater agreement was lowest for items on the Motivational subscale: 6 (*work*), 7 (*leisure activities*), 8 (*financial management*), 9

(*motivation for treatment*), as well as item 13 (*social network*). SAPROF total scores were analyzed and revealed very good consistency ($r = .86$, $p < .001$).

Table 6

Kappa Coefficient Scores for Intrarater and Interrater Reliability

SAPROF Item	Intrarater scores (*n*=13)	Interrater scores (*n*=8)
	Kappa Values	
1. Intelligence	.52**	.50*
2. Secure attachment in childhood	.63**	.29
3. Empathy	.52**	.46*
4. Coping	.54**	.43
5. Self-control	.54**	.33**
6. Work	.41**	.09
7. Leisure activities	.49	.08
8. Financial management	.57**	.08
9. Motivation for treatment	.76**	.05
10. Attitudes toward authority	.43*	.22
11. Life goals	.44**	.27
12. Medication	.68**	.20
13. Social network	.30	.05
14. Intimate relationship	.32*	.27
15. Professional care	.46**	.20
16. Living circumstances	.70**	.53
17. External control	1.00**	1.00**

**Rater agreement is significant at the .01 level. *Rater agreement is significant at the .05 level. *Note.* Poor agreement <0, slight agreement .00-.20, fair agreement .21-.40, moderate agreement .41-.60, substantial agreement .61-.80, almost perfect/perfect agreement .81-1.00 (Landis & Koch, 1977).

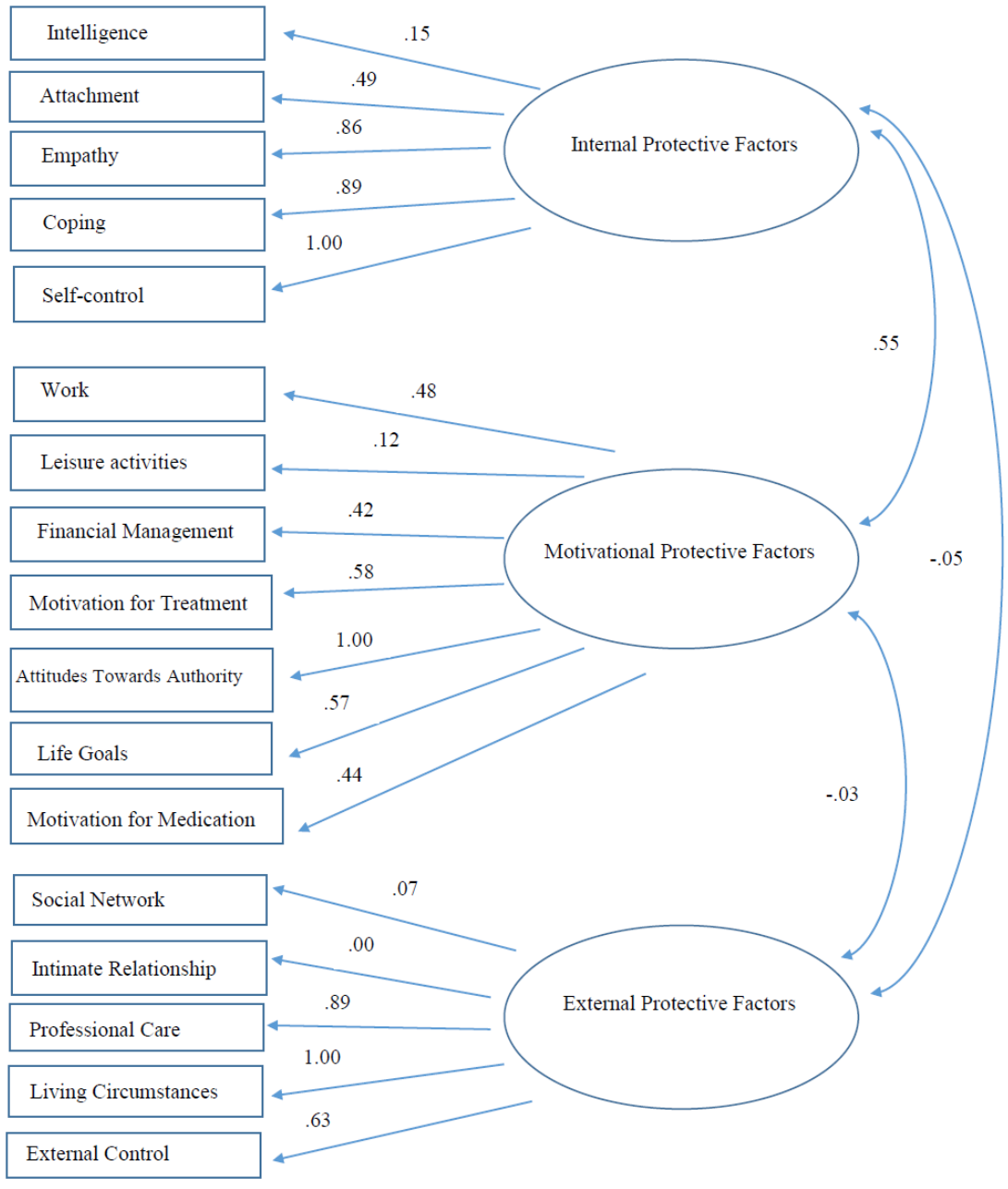
Confirmatory factor analysis. A confirmatory factor analysis (CFA) was performed to examine the factor structure of the SAPROF. The 3-factor structure was tested to determine whether it fit with the data, once constraints were imposed on the model based on a priori hypotheses. Results of the CFA model are presented in Figure 1, including the standardized beta coefficients for each item. Standardized beta coefficients represent the number of standard

deviations the dependent variable will change per increase in the standard deviation of the predictor variable (Field, 2009). The absolute fit was good, as assessed by the chi-square to degrees of freedom, $\chi^2/df = 1.28$. The RMSEA indicated a reasonable fit (RMSEA =.076, CI = 0.031, 0.11), and the PCLOSE= .14 suggests a close fit between the data and the hypothesized model (Brown, 2015; Browne & Cudeck, 1993; MacCallum et al., 1996). The CFI approached acceptable model fit (CFI=.85), whereas the NFI did not meet the acceptable cut-off (NFI=.58). Based on absolute and relative fit indicators, the data showed acceptable fit to the proposed model (Hooper, Coughlan, & Mullen, 2008; Hu & Bentler, 1999).

Of note, some factor loadings were relatively low compared to others (e.g., i*ntelligence* on the Internal subscale, *leisure activities* on the Motivational subscale, and both *social network* and *intimate relationship* on the External subscale). This suggests that some items within each factor are more related to each other than others for this particular sample. Nevertheless, the overall findings from the CFA suggest that the 3-factor model consisting of Internal, Motivational and External subscales, fit the sample data to an acceptable level.

Figure 1

3-Factor Confirmatory Factor Analysis for the SAPROF



SAPROF subscale internal reliability. Cronbach's α was calculated to assess the internal consistency of the SAPROF total score and subscales (Internal, Motivational and External). Cronbach's α is a function of the number of items on the scale and the average intercorrelation among the items (Cortina, 1993). In accordance with Kline (2000), the accepted cut-off for a reliable scale is .70. The reliability of the SAPROF (all items on the scale) met the accepted threshold, Cronbach's α =.72. The reliability of the Internal subscale approached the acceptable cut-off (α =.69). The reliability of the Motivational subscale was low (α =.59). In particular, items *leisure activities* (*r* =.13) and *financial management* (*r* =.10) did not correlate well with the Motivational subscale total score. The reliability of the External subscale also approached the acceptable cut-off, Cronbach's α =.65. Specifically, *social network* (*r* = .09) and *intimate relationship* (*r* = -.01) items did not correlate well with the External subscale total score. Overall, the Internal subscale appeared to be the most reliable SAPROF subscale, followed by the External and Motivational subscales.

Table 7 presents correlations among the SAPROF total score and subscale scores. Internal and Motivational subscales significantly correlated with one another; however, the External subscale did not correlate with the other two subscales. In order to increase the association between the External subscale and the other subscales, the last three items on the External subscale (*professional care, living circumstances, external control*) were removed, given that these items may be more representative of risk rather than protection. The External subscale still did not correlate with the Internal or Motivational subscales (*r* =.05, *p* =.71 and *r* = .25, *p* =.08), respectively.

Table 7

Correlations between SAPROF Total and Subscale Scores

	1	2	3	4
1. SAPROF Total score				
2. Internal subscale	.78**			
3. Motivational subscale	.85**	.58**		
4. External subscale	.41**	-.04	.05	

**p<.01

Hypothesis 1

The first research question addressed convergent validity of the SAPROF. Pearson's correlations were conducted between SAPROF scores and protective and risk factor measures (refer to Table 8). Specifically, the START Strength scale, a measure of protective factors, and the START Vulnerability scale, HCR-20 V3, LS/CMI and PCL-R, all measures of risk factors, were used to correlate with SAPROF total and subscale scores.

SAPROF total and subscale scores positively correlated with the Strength Scale of the START. The findings were mixed with respect to establishing convergent validity with risk measures. SAPROF total scores were negatively correlated with the Strength Vulnerability scale.

SAPROF total and subscale scores were significantly negatively correlated with HCR-20 V3 total and item scores, except for the External subscale of the SAPROF. SAPROF total and subscale scores were not correlated with LS/CMI total scores or individual subscales. One exception was the Procriminal Attitude/Orientation scale, which was significantly negatively correlated with the Internal Subscale, although this was a weak relationship. SAPROF total and subscale scores generally did not correlate with the PCL-R total or factor scores either. The Internal Subscale negatively correlated with PCL-R total scores and Factor 2 scores

(representatively of an antisocial lifestyle); however, this was a weak inverse relationship, and the effect size was small.

Spearman's correlations were used for ordinal scales, to determine convergent validity between the SAPROF judgement rating and the rating categories of the START, HCR-20 V3, and LS/CMI. The SAPROF judgement rating and the START Strength scale rating were positively correlated ($\rho = .67$, $p < .001$), reflective of a large effect size. Judgement ratings between the SAPROF and START Vulnerability scales were negatively correlated ($\rho = -.71$, $p < .001$), reflective of a very large effect size. Judgement ratings between the SAPROF and HCR-20 V3 revealed a negative association between the two measures ($\rho = -.51$, $p < .001$), indicative of a large effect size. SAPROF and LS/CMI judgement ratings were negatively correlated but not at a statistically significant level ($\rho = -.12$, $p = .48$), representative of a small effect size.

Overall, the results partially supported Hypothesis 1. Convergent validity was established for the SAPROF and START Strength scale, another protective factor measure, and was inversely related to the HCR-20 V3, a risk factor measure. Convergent validity was not established with the LS/CMI and PCL-R, two other risk measures, as expected. Correlations improved once measures were corrected for attenuation (the disattenuation score).

Table 8

Correlations between Protective Factor and Risk Factor Measures by SAPROF Subscale

Measure	SAPROF Total score	Disattenuation	Internal subscale	Motivational subscale	External subscale
Protective Factor					
START Strength scale	.84**	1.00	.79**	.81**	.04
Risk Factor					
START Vulnerability scale	-.83**	-1.00	-.80**	-.79**	-.03
HCR-20 V3					
Total score	-.52**	-.75	-.62**	-.40**	-.01
Historical Items	-.32*		-.42**	-.26	.05
Clinical Items	-.57**		-.65**	-.46**	-.01
Risk Management	-.35*		-.41**	-.23	-.06
Items					
LS/CMI					
Total score	-.14	-.20	-.19	-.09	.03
Criminal history Scale	.07		-.10	.21	.02
Education/employment Scale	.06		.02	.03	.06
Family/marital scale	.00		.00	.09	-.10
Leisure/recreation scale	-.13		.09	-.07	-.26
Companions scale	.00		-.07	.03	.06
Alcohol/drug problems scale	.04		-.07	-.01	.17
Procriminal attitude/orientation scale	-.29		-.35*	-.29	.14
Antisocial pattern scale	-.13		-.24	-.12	.13
PCL-R[1]					
Total score	-.21	-.27	-.31*	-.17	.01
Factor 1 Score	-.14		-.10	-.19	.03
Factor 2 Score	-.22		-.36*	-.14	.12

***$p < .001$. **$p < .01$. *$p < .05$. The reliability of the PCL-R total score was taken from Hare et al., 1990, in order to correct for attenuation.

Hypothesis 2

The second research question addressed divergent validity of the SAPROF. Correlations between the SAPROF, general functioning, and mental state measures were conducted.

Pearson's correlations were used to examine the relationship between SAPROF total and subscale scores with the GAF score (a measure of general functioning) and SANS/SAPS (mental state measure) (see Table 9).

The GAF was significantly correlated with the SAPROF total score, Internal subscale and Motivational subscale, and a large effect was established. The External subscale was not correlated with the GAF. The SANS score was negatively correlated with the SAPROF total score and subscale scores to a small degree, but not statistically significant, whereas the SAPS and the Composite SANS/SAPS score were both negatively and statistically correlated with the SAPROF total score, Internal and Motivational subscales.

Overall, the results revealed mixed findings with regards to the divergent validity of the SAPROF; therefore, Hypothesis 2 was partially supported. Across all subscales, the External subscale was the only SAPROF subscale to demonstrate divergent validity with general functioning and mental state measures. The SANS score also did not correlate with the SAPROF total and subscale scores as expected, suggesting that the SAPROF is not a measure of negative symptoms of schizophrenia. The GAF, SAPS and SANS/SAPS Composite score were found to correlate with SAPROF total scores and Internal and Motivational subscales. Correlations improved once measures were corrected for attenuation (the disattenuation score).

Table 9

Correlations between Mental State and Risk Measures by SAPROF Subscale

Measure	SAPROF Total score	Disattentuation	Internal subscale	Motivational subscale	External subscale
General Functioning					
GAF	.51**		.56**	.46**	-.02
Mental State					
SANS/SAPS					
SANS score	-.21	-.28	-.19	-.16	-.08
SAPS score	-.50**	-.66	-.41**	-.50**	-.07
SANS/SAPS Composite score	-.43**	-.56	-.41**	-.42**	-.03

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

Hypothesis 3

The predictive validity of SAPROF total and subscale scores was assessed for proxies of recidivism, which included medication administration, institutional misconduct and disposition breaches over the past year and six-month follow-up. Each proxy was assessed as both a dichotomous (presence or absence) and continuous variable (number of occurrences). Prior to conducting multiple linear and logistic regression analyses, both Pearson and Spearman's correlations were calculated (see Tables 10 and 11).

Table 10

Pearson Correlations between Continuous Predictors and Outcome Variables

Predictors and Outcomes	1	2	3	4	5	6	7	8
1. SAPROF total score								
2. SAPROF Internal subscale	.78**							
3. SAPROF Motivational subscale	.85**	.58**						
4. SAPROF External subscale	.41**	-.04	.05					
5. HCR-20 V3 total score	-.52**	-.62**	-.40**	-.01				
6. Number of PRN administrations	.09	.09	.01	.10	-.33			
7. Number of institutional misconducts	-.34*	-.44**	-.47**	.30*	.02	.25		
8. Number of disposition breaches	-.40**	-.43**	-.31*	-.06	-.28	-.15	.23	

**$p < .01$. *$p < .05$.

Table 11

Spearman Correlations between Continuous Predictors and Categorical Outcome Variables

Predictors and Outcomes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. SAPROF total score																
2. SAPROF Internal subscale	.79**															
3. SAPROF Motivational subscale	.79**	.55**														
4. SAPROF External subscale	.40**	-.02	.04													
5. HCR-20 V3 total score	.54**	-.63**	-.38**	-.03												
6. Index Score	.80**	-.77**	-.62**	-.21	.92**											
7. PRN administration[1]	-.04	-.33*	.00	.35*	.16	.12										
8. Institutional misconduct[1]	-.26	-.49**	-.32*	.28*	.37**	.36*	.23									
9. Disposition breaches[1]	-.42**	-.48**	-.30*	-.08	.35*	.43**	.10	.28*								
10. Privilege change[1]	.28	.22	.15	.21	-.29*	-.29*	.21	-.07	-.06							
11. Increase in privileges[1]	.36*	.31*	.30*	.13	-.43**	-.44**	.13	-.09	-.35*	.78*						
12. Security change[1]	.21	.15	.04	.17	-.24	-.23	.23	.10	-.07	.76**	.55**					
13. Decrease in security[1]	.34*	.27	.23	.10	-.41*	-.41**	.21	.05	-.40**	.56**	.79**	.75**				
14. PRN administration[2]	-.35*	-.61**	-.40**	.34*	.48**	.48**	.47**	.40**	.36*	-.03	-.17	.01	-.09			
15. Institutional misconduct[2]	-.24	-.54**	-.37*	.41**	.43**	.39**	.23	.50**	.24	.13	.05	.23	.11	.50**		
16. Disposition breaches[2]	.36*	-.36*	-.19	-.16	.24	.28	-.08	.19	.37*	-.16	-.22	.05	-.10	.24	.29*	

Note. [1]one year prior. [2] at six-month follow-up. **$p < .01$. *$p < .05$.

Medication administration. The administration of psychiatric PRN medication, was coded as either present or absent for medication in the past year. A logistic regression was conducted using SAPROF subscale scores as predictors. The Hosmer-Lemeshow test revealed a good fit of the model to the data, χ^2 (8) = 12.67, p =.12, and the test of the overall model was significant, χ^2 (3) = 10.81, p < .01 for SAPROF subscale scores (see Table 12). The proportion of variance accounted for by subscale scores was 30%, as measured by Nagelkerke's pseudo R^2. The results indicate that, as Internal subscale scores are higher, the odds of being administered a PRN were lower (OR = .67), and as External subscale scores are higher, the odds of being administered a PRN were higher (OR = 1.48).

Table 12

Logistic Regression for the Presence of PRN Administration for SAPROF Subscales

Predictor	B	SEβ	*P*	OR	95% CI OR
Internal subscale	-.41	.20	.04	0.67*	0.45, 0.98
Motivational subscale	.20	.17	.23	1.22	0.88, 1.71
External subscale	.40	.18	.03	1.48*	1.04, 2.12

Note. OR = Odds Ratio. *p < .05.

Multiple regression analysis was used to test whether SAPROF subscale scores predicted the total number of PRN administrations. Results indicated that neither the Internal, Motivational, or External subscale scores predicted the number of PRN administrations (β = -.01, p =.97, β = -.15, p =.55, β = .38, p =.06, respectively). The effect size was moderate, R=.42, R^2 =.17

PRN medication was further examined at a six-month follow-up. A logistic regression was conducted using SAPROF subscale scores as predictors. The Hosmer-Lemeshow test revealed a good fit of the model to the data, χ^2 (7) = 3.94, p = .79, and the test of the overall model was significant, χ^2 (3) = 32.37, p < .01 for SAPROF subscale scores (see Table 13). The proportion of variance accounted for by subscale scores was 69%, as measured by Nagelkerke's pseudo R^2. The

results signify that as Internal subscale scores are higher, the odds of being administered a PRN is lower (OR = .32), and as External subscale scores are higher, the odds of being administered a PRN is higher (OR = 2.29).

Table 13

Logistic Regression for the Presence of PRN Administration for SAPROF Subscales at Follow-up

Predictor	B	SEβ	*P*	OR	95% CI OR
Internal subscale	-1.15	.42	.01	0.32**	0.45, 0.98
Motivational subscale	-.22	.20	.26	0.80	0.88, 1.71
External subscale	.83	.30	.01	2.29**	1.04, 2.12

Note. OR = Odds Ratio. **$p < .01$.

Multiple regression analysis was used to test whether SAPROF subscale scores predicted the total number of PRN administrations at the six-month follow-up. The results indicated neither the Internal, Motivational, or External subscale scores predicted the number of PRN administrations ($\beta = -.04$, $p = .84$, $\beta = -.20$, $p = .30$, $\beta = .30$, $p = .06$, respectively). The effect size was medium, $R = .35$, $R^2 = .13$.

Institutional misconduct. Logistic regression was used to predict the presence or absence of institutional misconduct (both major and minor misconducts) over the past year for inpatients only (n=32). When SAPROF subscales were used as predictors, the Hosmer-Lemeshow test revealed a good fit of the model to the data, $\chi^2 (8) = 6.80$, $p = .56$, and the test of the overall model was significant, $\chi^2 (3) = 13.71$, $p < .01$ (refer to Table 14). Subscale scores accounted for 52% of the variance in the model, as measured by Nagelkerke's pseudo R^2. Specifically, the results indicated that the odds of committing institutional misconduct is lower (OR = .31) as Internal subscale scores are higher.

Table 14

Logistic Regression for the Presence of Institutional Misconduct for SAPROF Subscales

Predictor	B	SEβ	*p*	OR	95% CI OR
Internal subscale	-1.16	.48	.02	0.31*	0.12, 0.80
Motivational subscale	.36	.33	.27	1.43	0.76, 2.71
External subscale	-.23	.49	.64	0.79	0.30, 2.10

Note. OR = Odds Ratio. *p <.05.

Multiple regression analysis was used to test whether SAPROF subscale scores predicted the total number of institutional misconducts (both major and minor misconducts) among inpatients. Results indicated that neither Internal, Motivational, nor External subscale scores predicted the number of institutional misconducts (β = -.33, p =.17, β = -.26, p =.27, β = -.11, p=.51, respectively). The effect size was large, R =.58, R^2 = .33.

Institutional misconduct was further examined at a six-month follow-up in terms of the presence of institutional misconduct among inpatients (n=29). A logistic regression was conducted using SAPROF subscale scores as predictors. The Hosmer-Lemeshow test revealed a good fit of the model to the data, χ^2 (8) = 4.90, p = .77, and the test of the overall model was significant, χ^2 (3) = 18.33, p < .01 for SAPROF subscale scores (see Table 15). The proportion of variance accounted for by subscale scores was 67%, as measured by Nagelkerke's pseudo R^2. The results signify that as the Internal subscale score is higher, the odds of engaging in institutional misconduct is lower (OR = .29). These results were consistent with the findings from the previous year.

Table 15

Logistic Regression for the Presence of Institutional Misconduct for SAPROF Subscales at Follow-up

Predictor	B	SEβ	*P*	OR	95% CI OR
Internal subscale	-1.23	.53	.02	0.29*	0.10, 0.83
Motivational subscale	-.06	.26	.83	0.95	0.88, 1.58
External subscale	-.32	.67	.63	0.73	0.19, 2.72

Note. OR = Odds Ratio. *$p < .05$.

Multiple regression analysis was used to test whether SAPROF subscale scores predicted the total number of institutional misconducts (both major and minor misconducts) among inpatients at the six-month follow-up. Results indicated that Internal, Motivational, and External subscale scores failed to predict the number of institutional misconducts ($\beta = -.38$, $p = .15$, $\beta = -.07$, $p = .79$, $\beta = -.12$, $p = .50$, respectively). The effect size was large, $R = .48$, $R^2 = .22$.

Disposition breaches. The presence or absence of disposition breaches (both major and minor breaches) over the past year was predicted using logistic regression analyses. SAPROF subscales were assessed as predictors. The Hosmer-Lemeshow test revealed a good fit of the model to the data for subscale scores, $\chi^2(8) = 12.76$, $p = .12$, and the test of the overall model was significant, $\chi^2(3) = 12.36$, $p < .01$ (see Table 16). The proportion of variance in subscale scores was 29%, as measured by Nagelkerke's pseudo R^2. Specifically, Internal subscale scores accounted for this variance, which suggests that as Internal subscale scores increase, the odds of a disposition breach were lower (OR = .65).

Table 16

Logistic Regression for the Presence of Disposition Breaches by SAPROF Subscale

Predictor	B	SEβ	*p*	OR	95% CI OR
Internal subscale	-.43	.18	.02	0.65*	0.46, 0.93
Motivational subscale	-.08	.14	.56	0.92	0.70, 1.22
External subscale	-.06	.16	.71	0.94	0.69, 1.29

Note. OR = Odds Ratio. *$p < .05$.

Multiple regression analysis was used to test whether SAPROF subscale scores predicted the total number of disposition breaches (both major and minor breaches). The results of the regression indicated that three subscales explained 20% of the variance ($R^2 = .20$, $F (3, 46) = 3.80$, $p = .01$). As shown in Table 17, the Internal subscale significantly predicted the number of disposition breaches over the past year. Specifically, lower scores on the Internal subscale were more predictive of more institutional misconducts over the past year than higher scores on the subscale.

Table 17

Multiple Regression for Total Number of Disposition Breaches by SAPROF Subscale

Predictor	B	SEB	B
Internal subscale	-.21	.09	-.39*
Motivational subscale	-.04	.08	-.08
External subscale	-.05	.09	-.07

*$p < .05$.

Disposition breaches were further examined at a six-month follow-up, and a logistic regression was conducted using SAPROF subscale scores as predictors. The Hosmer-Lemeshow test revealed a good fit of the model to the data, $\chi^2 (7) = 2.72$, $p = .91$; however, the overall model only approached significance, $\chi^2 (3) = 7.07$, $p = .07$. Specifically, the Internal subscale accounted for this variance ($\beta = -.53$, $p = .05$). The proportion of variance accounted for by subscale scores

was 25%, as measured by Nagelkerke's pseudo R^2. Unlike the findings based on the previous year, SAPROF subscale scores did not predict disposition breaches at the six-month follow-up.

Multiple regression analysis was used to test whether SAPROF subscale scores predicted the total number of disposition breaches (both major and minor) at the six-month follow-up. The results of the regression indicated that the model was not significant (R^2 = .14, F (3, 46) = 2.25, p =.10). However, the Internal subscale scores predicted the number of disposition breaches at the six-month follow-up (β = -.37, p=.04), unlike the Motivational and External subscale (β = .06, p =.72, β = -1.05, p=.30), respectively. Therefore, the Internal subscale accounted for the variance in the model. Overall, SAPROF subscale scores predicted the number of disposition breaches over the past year but not over a six-month follow-up.

In summary, the SAPROF generally predicted proxies of recidivism as expected for Hypothesis 3. An alpha criterion of p <.01, was implemented and only results that met this pre-determined threshold were considered statistically significant. The SAPROF approached significance for predicting the presence of PRN administrations, but did not predict the total number of administrations over the past year. Similarly, the SAPROF predicted the presence but not the total number of PRN administrations upon a six-month follow-up. With regard to institutional misconducts, SAPROF predicted the presence of misconducts over the past year and upon a six-month follow-up among inpatients but not the number of misconducts at either time point. With respect to disposition breaches, the SAPROF was only predictive of the presence and total number of breaches over the past year, but not at a six-month follow-up.

Hypothesis 4

Hypothesis 4 concerned the predictive validity of the SAPROF in predicting risk management decisions over the last year. Specifically, a change in privilege and security level, and the direction of change was investigated.

Privilege level. A logistic regression was conducted using three SAPROF subscale scores as the predictors and privilege change (either yes or no) as the outcome. The Hosmer-Lemeshow test revealed a poor fit of the model to the data, χ^2 (8) = 16.27, p = .04, and the test of the overall model was not significant, χ^2 (3) = 5.57, p =.13. The proportion of variance in subscale scores was 14%, as measured by Nagelkerke's pseudo R^2. None of the SAPROF subscales predicted change in privilege level.

Multinomial regression analysis was conducted to further assess the relationship between SAPROF scores and change in privilege level. The direction of privilege level change over the past year was examined and the distribution was as follows: increase privileges (38.8%), decrease privileges (12.2%), and no change in privileges (49%). The no privilege change category was used as the base reference group.

The model was significant for SAPROF total scores, χ^2 (2) = 6.48, p = .04. The proportion of variance in total scores, as measured by the Nagelkerke's pseudo R^2 statistic, was 14%. As shown in Table 18, SAPROF total scores were significantly greater by a factor of 1.17 when privileges increased (p =.03) compared to individuals who did not receive a change in their privilege level. The findings for a decrease in privileges was not significant (p =.83). Conversely, the overall model was not significant for SAPROF subscale scores, χ^2 (82) = 94.40, p = .17. The proportion of variance in subscale scores, as measured by the Nagelkerke's pseudo R^2 statistic was 23%.

Table 18

Multinomial Logistic Regression of Direction of Privilege Level Change Predicted by SAPROF Total Scores

Predictor	B	SEβ	OR	95% CI OR	Wald Statistic
Decrease in privilege					
SAPROF Total score	-.02	.09	0.98	0.81, 1.19	.05
Increase in privilege					
SAPROF Total score	.15	.07	1.17*	1.02, 1.34	4.79

Note. OR = Odds Ratio. *$p < .05$

Security level. A logistic regression was conducted using SAPROF subscale scores as predictors and the presence or absence of change in security level as the outcome. The Hosmer-Lemeshow test revealed a poor fit of the model to the data, χ^2 (8) = 17.18, p = .03, and the test of the overall model was not significant, χ^2 (3) = 3.21, p = .36. The proportion of variance explained by subscale scores was 9%, as measured by Nagelkerke's pseudo R^2.

Multinomial regression analysis was conducted to further assess the relationship between SAPROF scores and change of security level. The direction of security level change over the past year was examined and the distribution was as follows: increase security/more restrictions (14.3%), decrease security/less restrictions (40.8%), and no change in security (44.9%). The no security change was used as the base reference group.

The model approached significance for SAPROF total scores, χ^2 (2) = 5.52, p = .06. Specifically, higher SAPROF total scores increased the likelihood that an individual would receive a decrease in security, however it was not statistically significant. The proportion of variance in total scores, as measured by the Nagelkerke's pseudo R^2 statistic, was 12%. The overall model was not significant for subscale scores either, χ^2 (6) = 8.87, p = .18. The proportion of variance in subscale scores, as measured by the Nagelkerke's pseudo R^2 statistic, was 19%.

Therefore, regardless of the direction of change in security level over the past year, neither subscale nor total SAPROF scores significantly predicted a change.

In sum, Hypothesis 4 was partially supported for the prediction of risk management decisions. An alpha criterion of $p < .01$, was implemented and only results that met this pre-determined threshold were considered statistically significant. SAPROF scores did not predict privilege change over the past year, however the increase of privilege level approached significance, suggesting that higher SAPROF scores may predict an increase in privileges. Although SAPROF scores did not predict security level change over the past year, the decrease of security level approached significance. This suggests that higher SAPROF scores may predict a lower level of security/fewer restrictions.

Hypothesis 5

The fifth research question addressed whether the SAPROF provides incremental predictive validity over and above risk assessment measures. This question was investigated using five dichotomous outcome variables: PRN administration, institutional misconduct, disposition breaches, increase in privilege level and decrease in security level (indicative of more freedoms and fewer restrictions). Hierarchical logistic regression analysis was conducted by entering HCR-20 V3 total scores in Step 1, followed by SAPROF total scores in Step 2.

Receiver Operating Characteristic (ROC) analysis was used to further examine the predictive validity for HCR-20 V3 total scores, SAPROF total scores and Index Scores (derived by subtracting SAPROF total scores from HCR-20 V3 total scores). To facilitate interpretation of the results, SAPROF total scores were reverse scored before they were entered into the model in order to correspond with the classification direction of the HCR-20 V3 total scores (e.g., higher

scores indicate greater risk level). For these analyses, higher scores on both measures were associated with a positive test (e.g., the presence of the outcome) for each ROC analysis.

Medication administration. The incremental predictive validity of SAPROF total scores over and above HCR-20 V3 total scores was examined to predict PRN administrations over the past year. The findings from hierarchical logistic regression analysis revealed that HCR-20 V3 total scores alone were not predictive of the presence of PRN administrations in Step 1, $\beta = .04$, $p = .34$. Further, in Step 2, SAPROF total scores did not add incremental predicative validity, $\beta = .05$, $p = .54$. The overall final model was not significant, $\chi^2 (2) = 1.30$, $p = .52$, indicating that neither HCR-20 V3 nor SAPROF total scores predicted whether a patient was administered a PRN over the past year.

ROC analyses were conducted to determine the predictive validity of the HCR-20 V3, SAPROF and Index Scores for the administration of a PRN, over the previous year (see Table 19). The Area Under the Curve (AUC) for the HCR-20 V3, SAPROF and Index Scores denote predictions that are just above chance and relate to a small effect size $d > .20$ (Rice & Harris, 2005). Differences between areas for all three AUC values were analyzed using pairwise comparisons using the DeLong, DeLong and Clarke-Pearson (1988) method. None of the differences were statistically significant. Table 21 presents the pairwise comparisons between all test variables.

The incremental predictive validity of the SAPROF was also measured at a six-month follow-up. The findings from hierarchical logistic regression analysis revealed that HCR-20 V3 total scores alone were significantly related to the presence of PRN administrations in Step 1, $\beta = .18$, $p = .01$. In Step 2, SAPROF total scores did not add incremental predictive validity, $\beta = -.07$, $p = .41$. Nevertheless, the overall final model was significant, $\chi^2 (2) = 12.27$, $p < .01$, indicating

that the shared variance of both the HCR-20 V3 and SAPROF total scores predicted the presence of PRN administrations better than the constant in the following six-months. When the SAPROF was added, the variance in the model only slightly increased, from 30% to 32%.

Further, ROC analyses were conducted to determine the predictive validity of the HCR-20 V3, SAPROF and Index Score for PRN administration over the six-month follow-up (see Table 20). The AUC values denote very large effect sizes for each measure, $d > .50$ (Rice & Harris, 2005). Differences between all three AUC values were analyzed and none was statistically significant (refer to Table 22).

Table 19

Area Under the Curve Values for HCR-20 V3, SAPROF and Index Scores for the Presence of Each Outcome over the Past Year

	AUC	SE	CI
PRN Administrations			
HCR-20 V3	.60	.09	[.44, .74]
SAPROF	.52	.09	[.37, .68]
Index Score	.57	.09	[.41, .72]
Institutional Misconduct[1]			
HCR-20 V3	.71	.13	[.46 .96]
SAPROF	.77*	.10	[.57, .97]
Index Score	.72	.12	[.50, .95]
Disposition Breaches			
HCR-20 V3	.70*	.07	[.56, .82]
SAPROF	.75**	.07	[.60, .86]
Index Score	.75**	.07	[.61, .86]
Privilege Level			
HCR-20 V3	.75**	.08	[.61, .90]
SAPROF	.72*	.08	[.56, .87]
Index Score	.76**	.08	[.60, .92]
Security Level			
HCR-20 V3	.73**	.07	[.59 .88]
SAPROF	.70*	.08	[.55, .85]
Index Score	.74**	.08	[.59, .89]

Note. AUC= Area Under the Curve. [1]$n = 32$. *$p < 0.05$. **$p < 0.01$.

Table 20

Area Under the Curve Values for HCR-20 V3, SAPROF and Index Scores for the Presence of Each Outcome at Six-month Follow-up

	AUC	SE	CI
PRN Administrations			
HCR-20 V3	.79***	.07	[.66, .92]
SAPROF	.71*	.08	[.56, .86]
Index Score	.79***	.07	[.66, .92]
Institutional Misconduct[1]			
HCR-20 V3	.41	.11	[.19 .63]
SAPROF	.37	.11	[.16, .58]
Index Score	.38	.11	[.17, .59]
Disposition Breaches			
HCR-20 V3	.69	.10	[.50, .89]
SAPROF	.79**	.07	[.66, .92]
Index Score	.72	.09	[.55, .90]

Note. AUC= Area Under the Curve. [1]$n = 29$ *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$

Table 21

Pairwise Comparisons of Area Under the Curve Values for HCR-20 V3, SAPROF and Index Scores by Outcome over the Past Year

Comparisons	Differences between AUC areas	SE	CI
PRN Administrations			
HCR-20 V3 vs. SAPROF	.07	.09	[-.09, .25]
SAPROF vs. Index Scores	.05	.06	[-.06, .16]
Index Score vs. HCR-20 V3	.03	.04	[-.05, .10]
Institutional Misconduct			
HCR-20 V3 vs. SAPROF	.06	.12	[-.17, .29]
SAPROF vs. Index Scores	.05	.08	[-.10, .20]
Index Score vs. HCR-20 V3	.01	.05	[-.08, .10]
Disposition Breaches			
HCR-20 V3 vs. SAPROF	.04	.08	[-.11, .20]
SAPROF vs. Index Scores	.00	.03	[-.02, .11]
Index Score vs. HCR-20 V3	.05	.03	[-.02, .11]
Privilege Level			
HCR-20 V3 vs. SAPROF	.04	.08	[-.11, .19]
SAPROF vs. Index Scores	.04	.05	[-.06, .14]
Index Score vs. HCR-20 V3	.00	.04	[-.07, .07]
Security Level			
HCR-20 V3 vs. SAPROF	.03	.08	[-.12, .19]
SAPROF vs. Index Scores	.04	.05	[-.06, .08]
Index Score vs. HCR-20 V3	.01	.04	[-.06, .08]

Note. AUC=Area Under the Curve.

Table 22

Pairwise Comparisons of Area Under the Curve Values for HCR-20 V3, SAPROF and Index Scores by Outcome at Six-month Follow-up

Comparisons	Differences between AUC areas	SE	CI
PRN Administrations			
HCR-20 V3 vs. SAPROF	.08	.08	[-.08, .24]
SAPROF vs. Index Scores	.08	.06	[-.04, .20]
Index Score vs. HCR-20 V3	.00	.03	[-.06, .06]
Institutional Misconduct			
HCR-20 V3 vs. SAPROF	.09	.12	[-.13, .32]
SAPROF vs. Index Scores	.03	.06	[-.09, .16]
Index Score vs. HCR-20 V3	.06	.06	[-.05, .17]
Disposition Breaches			
HCR-20 V3 vs. SAPROF	.10	.10	[-.09, .29]
SAPROF vs. Index Scores	.07	.07	[-.07, .21]
Index Score vs. HCR-20 V3	.03	.04	[-.04, .10]

Note. AUC=Area Under the Curve.

Institutional misconduct. The incremental predictive validity of SAPROF total scores over and above HCR-20 V3 total scores was examined to predict institutional misconduct over the past year. This calculation included inpatients only (n = 32). The findings from hierarchical logistic regression analyses revealed that HCR-20 V3 total scores alone did not significantly predict institutional misconduct in Step 1, $\beta = .11$, $p = .10$. Further, in Step 2, SAPROF total scores did not add incremental predictive validity but approached significance, $\beta = -.22$, $p = .07$ (see Table 23). Although neither the HCR-20 V3 nor SAPROF total scores predicted whether an inpatient engaged in institutional misconduct over the past year, the overall final model was significant. This suggests that the shared variance of the HCR-20 V3 and SAPROF total scores predicted greater institutional misconduct better than just the constant, $\chi^2 (2) = 6.84$, $p = .03$, which accounted for the significant model. When the SAPROF was added, the variance in the model increased from 13% to 29%.

Table 23

Incremental Predictive Validity of SAPROF Total Scores in Predicting the Presence of Institutional Misconduct over the Past Year

	β	SEβ	*p*	OR	95% CI OR	Nagelkerke's R^2
Step 1						
HCR-20 V3	.02	.07	.77	1.02	0.89, 1.18	.13
Step 2						
SAPROF	-.22	.12	.07	0.81	0.64, 1.02	.29

Note. OR = Odds Ratio. Final model χ^2 (2) = 6.84, $p < .05$.

ROC analyses were conducted to determine the predictive validity of the HCR-20 V3, SAPROF and Index Scores for the presence of institutional misconduct over the past year (see Table 19). The AUC for the SAPROF, HCR-20 V3, and Index Score represents a very large effect size, $d > .50$ (Rice & Harris, 2005). Of note, the SAPROF was the only measure that was statistically significant. Differences between all three AUC values were analyzed and none were statistically significant (refer to Table 21).

The incremental predictive validity of the SAPROF was also measured at a six-month follow-up for institutional misconduct. Only inpatients at the six-month follow-up were included in these analyses (n = 29). The findings from hierarchical logistic regression analysis revealed that HCR-20 V3 Total scores alone were not significantly related to the presence of institutional misconduct in Step 1, β = .16, p =.06. In Step 2, SAPROF scores added incremental predictive validity β = -.34, p = .04 (see Table 24). The overall final model was significant, χ^2 (2) = 10.23, $p < .01$, indicating that SAPROF scores predicted institutional misconduct in the six-month follow-up, over and above the HCR-20 alone. Once the SAPROF was added to the model, it accounted for an increase in variance from 24% to 48%.

Further, ROC analyses were conducted to determine the predictive validity of the HCR-20 V3, SAPROF and Index Score for institutional misconduct over the six-month follow-up (see Table 20). The AUC values denote medium effects for each measure, none statistically significant (Rice & Harris, 2005). Differences between all three AUC values were analyzed and none were statistically significant (refer to Table 22).

Table 24

Incremental Predictive Validity of SAPROF Total Scores in Predicting the Presence of Institutional Misconduct over a Six-month Follow-up

	β	SEβ	*p*	OR	95% CI OR	Nagelkerke's R^2
Step 1						
HCR-20 V3	.03	.09	.74	1.03	0.86, 1.24	.24
Step 2						
SAPROF	-.34	.17	.04	0.71	0.51, 0.99	.48

Note. OR = Odds Ratio. Final model χ^2 (2) = 10.23, $p < .01$.

Disposition breaches. The incremental predictive validity of SAPROF total scores over and above HCR-20 V3 total scores was examined to predict the presence of disposition breaches over the past year. The findings from hierarchical logistic regression analysis revealed that HCR-20 V3 total scores alone were found to be significantly related to an increase in the presence of disposition breaches in Step 1, β = .11, $p = .02$. In Step 2, SAPROF total scores added incremental predictive validity, β = -.16, $p = .04$. The final model was significant, indicating that both HCR-20 V3 and SAPROF total scores predicted a disposition breach significantly better than just the constant, χ^2 (2) = 11.01, $p < .01$ (see Table 25). These results suggest that the model improved, compared to just HCR-20 V3 total scores alone, χ^2 (1) = 6.52, $p = .01$. Further, when the SAPROF was added to the model, the variance accounted increased from 16% to 27%.

As HCR-20 V3 total scores become higher, the odds of a patient breaching their disposition are also higher (OR = 1.07), although not significantly. Further, the results indicate

that as SAPROF total scores become higher, the odds of a patient breaching their disposition will significantly lower (OR = .85), suggesting that more protective factors are predictive of fewer disposition breaches over the past year.

Table 25

Incremental Predictive Validity of SAPROF Total Scores in Predicting the Presence of Disposition Breaches

	β	SEβ	*p*	OR	95% CI OR	Nagelkerke's R^2
Step 1						
HCR-20 V3	.06	.05	.22	1.07	0.96, 1.18	.16
Step 2						
SAPROF	-.16	.08	.04	0.85*	0.73, 1.00	.27

Note. OR=Odds Ratio. Final model χ^2 (2) = 11.02, p <.01. *p < .05.

ROC analyses were conducted to determine the predictive validity of the HCR-20 V3, SAPROF and Index Scores for the presence of disposition breaches over the past year (refer to Figure 2 and Table 19). These AUC values for the HCR-20 V3 represents moderate effect, whereas the SAPROF and Index Scores represent a large effect size d > .80 (Rice & Harris, 2005). Differences between all AUC values were analyzed and none were statistically significant (see Table 21).

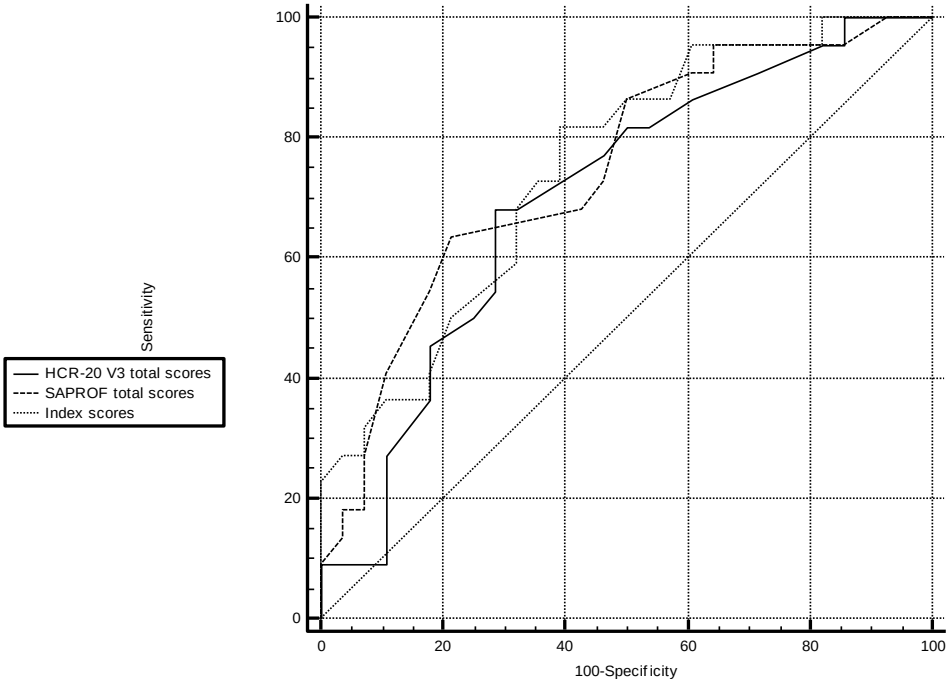
Additional analyses were performed by dividing the outcome variable into major (elopements, harm towards others, possession of weapons) and minor (substance use, refusal of medication, absent at community residence) disposition breaches. Base rates were approximately equal for minor breaches (n = 15, 30%) and major breaches (n =13, 26%) over the previous year. The SAPROF predicted minor disposition breaches but not major disposition breaches (overall model χ^2 (1) = 6.32, p <.01, Nagelkerke's pseudo R^2 =.17). Upon closer examination, substance use accounted for the majority of minor breaches, and SAPROF total scores distinguished individuals from the total sample who breached their disposition using substances, from those

who did not ($t(26) = -2.20$, $p = .03$). Therefore, individuals who breached their disposition by using substances tended to have lower SAPROF scores than those who did not.

The incremental predictive validity of the SAPROF was also measured at a six-month follow-up for disposition breaches. The findings from hierarchical logistic regression analysis revealed that HCR-20 V3 total scores alone did not significantly predict the disposition breaches in Step 1, $\beta = .10$, $p = .13$. Further Step 2, SAPROF scores did not add incremental predictive validity $\beta = -.16$, $p = .14$. The overall final model was not significant, $\chi^2 (2) = 5.03$, $p = .08$, indicating that neither the HCR-20 V3 nor SAPROF combined predicted disposition breaches in the six-month follow-up. Nevertheless, the model did improve and once the SAPROF was added, it accounted for an increase in variance from 9% to 18%.

Figure 2

ROC Graph of Test Variables Predicting the Presence of Disposition Breaches



Privilege level. The incremental predictive validity of SAPROF total scores over and above HCR-20 V3 total scores was examined as they relate to an increase in privilege level.

The findings from hierarchical logistic regression analysis revealed that HCR-20 V3 total scores alone significantly predicted an increase in privilege level in Step 1, β= -.14, $p < .01$. In Step 2, SAPROF total scores did not add incremental predictive validity, β = .09, $p = .26$ (see Table 26). Nevertheless, the overall final model was significant χ^2 (2) = 10.58, $p < .01$, indicating that both the HCR-20 V3 and SAPROF combined accounted for the slight increase in variance.

Table 26

Incremental Predictive Validity of SAPROF Total Scores for Predicting an Increase in Privilege Level

	β	SEβ	*p*	OR	95% CI OR	Nagelkerke's R^2
Step 1						
HCR-20 V3	-.14	.05	.01	0.87**	0.79, 0.96	.23
Step 2						
SAPROF	.09	.08	.26	1.10	0.94, 1.27	.26

Note. OR= Odds Ratio. Final model χ^2 (2) = 10.58, $p < .01$ **$p<.01$.

ROC analyses were conducted to determine the predictive validity of the HCR-20 V3, SAPROF and Index Scores for the increase in privilege level (refer to Table 19). The AUC's for HCR-20 V3, SAPROF and Index Score denote predictions slightly above chance and relate to medium-small to small effect sizes (Rice & Harris, 2005). Differences between all three AUC values were analyzed and none were statistically significant (see Table 21).

Security level. The incremental predictive validity of SAPROF total scores over and above HCR-20 V3 total scores was examined to predict a decrease in security level. The hierarchical logistic regression analysis revealed that HCR-20 V3 total scores alone were significantly related to a decrease in security level in Step 1, β = -.13, $p < .01$. In Step 2, SAPROF total scores did not add incremental predictive validity, β = .07, $p = .36$ (refer to Table 27). However, the overall final model was significant, indicating that both HCR-20 V3 and SAPROF total scores predicted a decrease in security level χ^2 (2) = 9.38, $p < .01$. This was also the case for HCR-20 V3 total scores alone, χ^2 (1) = 8.52, $p < .01$, suggesting that HCR-20 V3 total scores accounted for the variance.

Table 27

Incremental Predictive Validity of SAPROF Total Scores for Predicting a Decrease in Security Level

	β	SEβ	*P*	OR	95% CI OR	Nagelkerke's R^2
Step 1						
HCR-20 V3	-.13	.05	.01	0.88**	0.80, 0.97	.22
Step 2						
SAPROF	.07	.08	.36	1.07	0.92, 1.24	.24

Note. OR=Odds Ratio. Final model χ^2 (2) = 9.38, $p < .01$. **$p<.01$.

ROC analyses were conducted to determine the predictive validity of the HCR-20 V3, SAPROF and Index Score for the decrease in security level (see Table 19). The AUC for the HCR-20 V3, SAPROF, and Index Score correspond to predictions that are just above chance and relate to a small to moderate and small effect sizes (Rice & Harris, 2005). Differences between all three AUC values were analyzed and none were statistically significant (see Table 21).

In summary, Hypothesis 5 was partially supported, as the SAPROF incrementally predicted several of the outcome variables. An alpha criterion of $p < .01$, was implemented and only results that met this pre-determined threshold were considered statistically significant. The SAPROF did not add incremental predictive validity for PRN medication at any time point. The SAPROF added incremental predictive validity for institutional misconduct over a six-month follow-up, although this finding only approached significance, but not over the previous year. SAPROF scores approached significance for adding incremental predictive validity for the presence of disposition breaches, over and above HCR-20 V3 total scores over the previous year, but not at the six-month follow-up. The SAPROF did not added incremental predictive validity for privilege or security level change. ROC analyses determined that the SAPROF, HCR-20 V3 and Index Score predicted all outcome variables over the last year, with large to very large effects. Further,

ROC analyses demonstrated very large effect sizes for the SAPROF, HCR-20 V3 and Index Score for PRN administration and disposition breaches, and medium effects for institutional misconducts over the six-month follow-up. No differences were found between measures using pairwise comparisons, indicating that differences ROC values were not statistically significant across all outcome variables.

Hypothesis 6

Analyses for Hypothesis 6 were exploratory and focussed on a variety of patient characteristics. Three of the six patient characteristics were continuous: therapy (number of individual and group therapy sessions attended), institutionalization (the length of time in years residing at Ontario Shores), and disposition (length of time in years residing on current disposition). The remaining three patient characteristics were dichotomous: patient status (inpatient or outpatient), nature of index offence (violent or nonviolent index offence), and vocational activities (participated in vocational activities inside or outside the hospital or not).

SAPROF total scores and group differences were calculated for patient characteristics (refer to Table 28). Median-splits were computed for all continuous patient characteristics.

Table 28

SAPROF Mean Scores and Group Differences for Patient Characteristics

Patient Characteristics		SAPROF Scores	Group Differences
Patient status	Inpatients	*M*=19.69 (5.46)	
	Outpatients	*M*= 17.22 (4.28)	*t*(48) = -1.77
Index offence	Violent	*M*=18.94 (5.37)	
	Nonviolent	*M*=18.50 (4.86)	*t*(48) = .28
Therapy sessions	High (6 or more sessions)	*M*=21.00 (4.92)	
	Low (less than 6 sessions)	*M*=16.29 (4.56)	*t*(45) = 3.37***
Vocational activities	Participated	*M*=21.79 (4.69)	
	Did not participate	*M*=17.64 (4.92)	*t*(48) = 2.71**
Length on disposition	High (more than 1.75 years)	*M*=18.04 (5.72)	
	Low (less than 1.75 years)	*M*=19.56 (4.53)	*t*(48) = -1.04
Length of institutionalization	High (more than 3.5 years)	*M*=17.40 (4.80)	
	Low (less than 3.5 years	*M*=20.20 (5.23)	*t*(48) = -1.97*

Note. Median-splits were computed for continuous variables. **p*=.05, ***p*=.01, ****p*=.001

Given the number of patient characteristics collected and there were no theoretical assumptions for which variables may be most important, a stepwise logistic regression was conducted. Stepwise regression analyses are often used in the exploratory stages of model building in order to determine which predictors are most useful (Tabachnik & Fidell, 2000; Thayer, 2002). All predictor variables (SAPROF total scores and six patient characteristics) were entered in Step 1. In Step 2, all interaction terms were entered. A forward conditional method was used, which involves starting without any variables in the model, testing each additional variable using the selected criterion ($p < .05$), and adding the variable that improves the model most.

Prior to creating interaction terms, all continuous predictor variables were centred by subtracting the mean from each individual score in order to reduce multicollinearity. Interaction terms were created by multiplying centred SAPROF total scores with centred patient variables. A

stepwise logistic regression was run for each of the five outcome variables: presence of PRN administration, presence of institutional misconduct, presence of disposition breaches, increase in privilege level and decrease in security level, over the past year.

The findings revealed that interaction terms were not statistically significant, which supports Hypothesis 6. Therefore, patient characteristics did not influence or moderate the relationship between SAPROF scores and the outcomes of interest.

Discussion

The aim of the study was to examine the psychometric properties of the SAPROF among a sample of forensic patients. All previous studies on the SAPROF used archival data to assess the reliability and validity. The current study built upon previous research and included a sample of 50 NCRMD patients from an Ontario hospital, using a prospective and retrospective research design. Overall, evidence in support or partial support of the study hypotheses were found, thereby contributing to the risk/protective factor literature with a forensic population. Results revealed good reliability, construct validity and predictive validity for the SAPROF. The generalizability of the study sample will be addressed below, followed by a discussion of the findings for study hypotheses, study limitations, and clinical and future research implications.

Sample Characteristics

The characteristics of the study sample differed from previous studies examining the SAPROF. SAPROF total scores ranged in the literature from *M*= 10.93 (*SD*= 5.10) (Turner et al., 2014), to *M*= 23.35 (*SD*= 5.25) (Doyle et al., 2014). The mean SAPROF score in the current study was *M*=18.80 (*SD*= 5.16), which is approximately mid-way from the lowest and highest reported scores. Differences between the current sample and other study samples may have accounted for the variation in mean total scores.

Further, differences were noted between the current study and previous studies using Dutch samples. The primary psychiatric diagnosis in the current study was Schizophrenia or a Psychotic Spectrum Disorder (92%), whereas other studies reported that only 15% -19% of the sample had either a psychotic disorder (de Vries Robbe et al., 2015) or an Axis I disorder, more generally (de Vries Robbe et al., 2011). The prevalence of substance abuse or dependence disorders for the

current study was comparable to past studies; for example, de Vries Robbe and colleagues (2015) and de Vries Robbe and colleagues (2011) reported rates of 65% and 68% respectively.

Compared to other studies, the prevalence of personality disorders/traits in the current study was nearly half that reported in de Vries Robbe and colleagues (2015 and 2011). Given the higher rates of personality disorders/traits in previous studies, specifically cluster B which is marked by dramatic and impulsive behaviour, it may be expected that the average SAPROF total scores would be lower in previous samples, compared to the current study. Although SAPROF scores were relatively high in the present study, there was a high rate of prior offending among the current sample. In particular, 88% had a previous offence and 40% had a juvenile record, suggestive of long-standing criminal behaviour.

The sample demographics of the current study are similar to those reported by Martin and Martin (2016) and Reimann and Nussbaum (2011); these studies also used a sample of forensic patients from Ontario Shores. Although the current sample only included a portion of the total forensic population at the hospital (N=250), the current findings may be considered representative of all forensic patients at the hospital, given the similarity in characteristics to past studies (e.g., patient age, sex, diagnoses). Moreover, the sample demographics of the current study resembles a large scale study of NCR patients in Canada from 2000 – 2005 (Crocker et al., 2015). A report by Crocker and colleagues (2015) revealed that 94% of NCR individuals were diagnosed with a severe mental illness, 84% were males, and the average age was 41 years old. Differences in the present sample included higher prevalence rates of substance use disorders and personality disorders, and the nature of the index offence was generally more serious in nature (e.g., assault, aggravated assault, murder and attempted murder). Taken together, the current sample is representative of Ontario forensic patients and NCR patients across Canada.

Reliability

The SAPROF was found to be a reliable measure. Items on the Internal subscale were generally most reliable, as demonstrated by intrarater administrations. Internal subscale items are arguably more difficult to score on the SAPROF than Motivational and External subscale items, given that they rely on abstract information (e.g., *secure attachment in childhood*), as opposed to concrete information (e.g., *financial management*). As a result, it is possible that more information was gathered to score Internal subscale items in order to compensate, resulting in higher reliability scores across administrations.

External control was the most reliable item on the SAPROF, as determined by perfect agreement by both intra- and interrater administrations. This item was relatively simple to score because it was coded based on whether the individual was an inpatient or outpatient. *Self-control* was another item that demonstrated good reliability across intra-rater and interrater administrations, showing fair to moderate agreement. Self-control has been shown to be an important variable in predicting violent behaviour, and thus both a valid and reliable protective factor (Coid et al., 2015). SAPROF total scores demonstrated almost perfect agreement between two raters. This suggests that although there was poor reliability among some items, the SAPROF is a very reliable measure overall, which is more clinically relevant than the reliability of individual items.

Analyses revealed stronger intrarater than interrater reliability across SAPROF items and SAPROF total scores, which was to be expected. Methodological and individual differences between raters (e.g., the author and a clinician) may have accounted for this discrepancy. Of note, the average time between administrations of the SAPROF for intrarater reliability was six months. A researcher conducted intrarater reliability, whereas a clinician conducted the interrater

reliability. The discrepancy between researcher and clinician ratings is well-documented in the literature. In previous studies, researchers tended to yield higher risk estimates and may be more likely to adhere to coding rules than clinicians (Penney et al., 2014; Vincent, Guy, Fusco, & Gershenson, 2011). In contrast, clinicians frequently rely on their relationship with the patient when gathering information for coding. Similarly, clinicians may attempt to prove their personal predictions as invalid because they may be directly involved in the care of the patient and have a vested interest in the risk estimate assigned (de Vogel & de Ruiter, 2004; Vojt, Thomson, & Marshall, 2013). However, this may not be as relevant to the current study, as the second rater was blind to the patient being scored.

Other differences between coders are noteworthy, as well. The interrater coder had more clinical experience than the intrarater coder, which could have influenced ratings. Further, the interrater coder did not meet with participants in person and relied solely on scoring sheets and audio recordings of the interview. Therefore, the interrater coder was unable to incorporate the individual's clinical presentation or rely on nonverbal cues of communication for scoring, which may have influenced the accuracy of the reliability scores. Despite the disparities between raters, the scoring of the SAPROF simulated the methodology used in a clinical context (e.g., using in-person interviews and file information to score items and conducting interviews on patient units). This is essential, given that determining the reliability of a measure involves examining how it fares in real-world conditions in order to achieve ecological validity (Kazdin, 2003; Penney et al., 2014).

Construct Validity

Construct validity is the degree to which a test measures what it claims to be measuring (Kazdin, 2003). A confirmatory factor analysis (CFA) was conducted to verify the factor structure

of the SAPROF by testing the relationship between items and subscales. Some items on the SAPROF did not fit as well to the model for this particular sample (e.g., *intelligence*, *financial management*, *leisure activities*, *social network* and *intimate relationship*). This may have been due to the restricted variance in the sample, for example, only one individual reported a stable intimate relationship. Nevertheless, the CFA revealed an acceptable fit between the current data and the proposed model. Thus, the CFA supported the 3-factor model, consisting of Internal, Motivational and External subscales (de Vogel et al., 2011).

Convergent and divergent validity are two forms of construct validity that were examined in Hypotheses 1 and 2 of the present study. A strong positive relationship between the SAPROF and Strength Scale of the START was established, which is indicative of both measures tapping into a similar construct. This is likely due to the overlap between approximately six items (e.g., occupational/work, recreational/leisure activities, social support/social network, plans/life goals, coping, and material resources/financial management). The similarity in items suggests that both measures are tapping into many of the same components of protection.

Similarly, a strong negative relationship was found between the Vulnerability Scale of the START, the HCR-20 V3, and the SAPROF, which adds further support for the notion that protective factors may represent the reverse of risk factors (Webster et al., 2004). Abidin and colleagues (2013) also demonstrated convergent validity between the SAPROF, START and HCR-20 version 2. This may further suggest that risk and protection exist on the same continuum; however, because a perfect positive or negative relationship was not found, the measures may be tapping into somewhat distinct constructs.

Convergent validity for the SAPROF was not fully supported by the study findings. Contrary to expectation, there was little to no relationship between the SAPROF, LS/CMI, and

PCL-R. Post hoc analyses determined that all three risk measures (LS/CMI, PCL-R and HCR-20 V3) were positively correlated with one another at a significant level. Although the SAPROF was negatively correlated with the HCR-20 V3, convergent validity was not established for the LS/CMI or PCL-R. One possible explanation for this finding may be that the LS/CMI and PCL-R are tapping into different aspects of a common construct than the HCR-20 V3. For example, risk factors on the LS/CMI and PCL-R are largely based on antisocial attitudes and behaviours to predict risk for correctional samples (Walters, 2011), whereas risk factors on the HCR-20 V3 (e.g., clinical and risk management scales) may be more inclusive of psychiatric symptomology to assess risk and validated on mentally disordered offenders (Grann, Belfrage, & Tengstrom, 2000). Further, in Yoon et al.'s study (2011) investigating sexual offending, the SAPROF was associated with the SVR-20 but not the Static-99. This may suggest that the SAPROF better corresponds with instruments consisting of dynamic items or SPJ tools rather than static items or actuarial measures.

Divergent validity was partially established for the SAPROF. The SAPROF did not correlate with the SANS, indicating that these measures are tapping into different constructs. A strong positive relationship was found between the SAPROF and the GAF, and a strong negative relationship was found between the SAPROF and SAPS. Therefore, protective factors, positive symptoms of schizophrenia, and general functioning seem to be related. These findings make sense given that the primary reason for forensic patients' hospitalization is due to positive symptoms of psychosis (e.g., delusions and hallucinations), which is related to their offending behaviour. Therefore, it might be expected that a greater number of positive symptoms is associated with less protection.

Abidin and colleagues (2013) also found a strong positive association between the SAPROF and GAF. The GAF performed as well as many assessment instruments suggesting that global functioning may be an important indicator of risk and protection (Abidin et al., 2013). This study also found a strong inverse relationship between the SAPROF and both positive and negative symptoms of schizophrenia, using the Positive and Negative Syndrome Scale (PANSS; Kay, Fiszbein, & Opler, 1987). Nevertheless, the current finding suggests that the SAPROF is unrelated to negative symptoms of schizophrenia (e.g., lack of motivation, sociality, speech, expression, etc.).

Once disattenuated scores were calculated, correlations between the SAPROF and other measures (e.g., START, HCR-20 V3, LS/CMI, PCL-R, SAPS, SANS) were even stronger. Given that measurement error weakens associations between measures, accounting for the reliability of each measure resulted in greater accuracy of correlation estimates between the SAPROF and other measures.

Predictive Validity

The outcome variables investigated in the SAPROF literature generally focus on criminal offending, specifically, violent and sexual recidivism (de Vries Robbe et al., 2015). In contrast, the present study focussed on predicting risk management decisions, that is, changes in level of privilege and security, and short-term outcomes that may serve as proxies of criminal recidivism, (e.g., the administration of PRN medication, institutional misconduct and disposition breaches). These outcome variables are particularly important considering that ORB decisions have a direct impact on where the patient should reside, as well as the conditions necessary for effectively managing patient risk.

PRN administration. The SAPROF demonstrated good predictive validity for PRN administration at six-month follow-up, in accordance with Hypothesis 3. Although the SAPROF predicted whether a patient was administered a PRN over the past year, this finding approached significance only once the alpha criterion was applied. Higher scores on the SAPROF lowered the likelihood that a patient was administered a PRN. Upon further examination, a higher Internal subscale score was associated with lower odds of being administered a PRN, whereas a higher External subscale scores was associated with higher odds of being administered a PRN. This finding was consistent over both time points. One possible explanation for this finding is that higher scores on the Internal subscale suggests that the individual possesses the necessary internal characteristics, such as self-control and coping skills, to manage their thoughts, emotions, and behaviour. As such, they may not require PRN medication to sedate or dampen problematic thoughts, emotions and behaviours.

Although the association between protective factors and PRN administration among forensic patients has not been widely investigated in the literature, Hales and Gudjonsson (2014) examined demographic characteristics in relation to PRN medication. Among patients residing on a medium secure forensic unit, patient age was found to predict PRN administration. Particularly patients younger than the mean (*M*= 33 years old). It is possible that with older age and maturation, individuals possess more protective factors such as coping skills and self-control, serving to manage their psychiatric symptoms and behaviours.

In the current sample, few individuals received a PRN for aggression, anxiety, or sleep difficulties. More likely, patients were prescribed a scheduled antipsychotic medication, anxiolytics, and antidepressant medication. Receiving scheduled medication may have confounded the study findings, making it difficult to determine which medication was managing

symptoms and behaviour. Nevertheless, the current study found that patients with greater internal characteristics were less likely to need PRN medication, whereas individuals receiving more external forms of control required more PRN medication. This finding is consistent with descriptions of the External subscale in the literature (de Vogel, et al., 2011; de Vries Robbe et al., 2011). Unlike other subscales on the SAPROF, higher scores on the External subscale is associated with higher levels of supervision necessary to manage the individual's risk. Therefore, it would be expected that more PRN medication is required for individuals deemed higher risk.

Misconducts. Institutional misconduct comprised several rule breaking and aggressive behaviours occurring on the inpatient unit. SAPROF scores predicted institutional misconduct over the previous year and at the six-month follow-up. Similarly, both the presence and total number of disposition breaches were predicted by the SAPROF over the previous year. Abdin and colleagues (2013) reported that the SAPROF predicted inpatient violence among forensic mental health patients, compared to other protective and risk measures (e.g., HCR-20, PANSS, S-RAMM, and DUNDRUM). de Vries Robbe and colleagues (2016) also found that the SAPROF predicted aggressive incidents during hospitalization, specifically scores on the Motivational subscale.

Additionally, Braithwaite and colleagues (2010) and Chu and colleagues (2011) found that the START predicted aggression towards others, unauthorized leave, substance use while hospitalized, interpersonal violence, verbal threats and any inpatient aggression, upon a short-term follow-up of one month. This adds further support for the finding that protective factor measures can predict forensic patient behaviour, which is useful information for clinicians and review boards when making decisions about managing risk.

Incremental Predictive Validity

An important finding of the study was that the SAPROF approached significance for incrementally predicting institutional misconducts over the six-month follow-up and disposition breaches over the previous year. As expected, higher scores on the HCR-20 V3 and lower scores on the SAPROF were predictive of breaches and misconducts. The effect size doubled once the SAPROF was added to the HCR-20 V3. This suggests that with enough power, a significant finding may have been established. A potential explanation for why disposition breaches were not predicted upon follow-up may be due to short time lengths investigated. It is possible that with more time, more disposition breaches will be committed by patients. When compared to the HCR-20 V3 and Index Score, the SAPROF was the only measure that significantly predicted disposition breaches at follow-up, although pairwise comparisons were not statistically significant.

Disposition breaches were not predicted at follow-up likely due to the low base rates, making it difficult detect a significant association. Few patients breached their disposition one or more times at the six-month follow-up (n = 7, 14%), whereas the distribution of breaches over the previous year appear to be more evenly distributed across the sample. It is also possible that patients who breached their disposition over the past year were transferred to higher levels of supervision, preventing them from breaching their conditions over the six-month follow-up. Specifically, there were three cases in which the patient was transferred to maximum security psychiatric hospital/prison and disposition breaches at follow-up could not be tallied and were not included in analyses.

Further analyses revealed that the SAPROF predicted minor breaches but not major breaches. Most minor breaches were due to substance use. Individuals who were caught using

substances had lower SAPROF scores than individuals who were not caught/used substances. Thus, the SAPROF was able to discriminate individuals who committed a specific type of disposition breach, adding to its clinical utility. Substance use among NCRMD persons is a growing concern, given that the number of NCRMD persons diagnosed with a psychotic spectrum disorder comorbid with substance abuse has gradually increased over time (Simpson et al., 2014).

Based on increases in effect size, protective factors provided unique information when combined with risk factors to predict certain types of patient behaviour. Thus, in order to decrease the likelihood of committing a disposition breach or institutional misconduct, both the increase of protective factors and decrease of risk factors (e.g., on the HCR-20 V3) are equally important. This suggests that greater statistical power and a larger sample may have increased the likelihood of detecting a significant result.

Three previous SAPROF validation studies examined the incremental predictive validity of the SAPROF. Index scores were used, subtracting the SAPROF total score from both the HCR-20 and SVR-20, to predict desistance from violent and sexual reoffending respectively. de Vries Robbe and colleagues (2011, 2013, 2015) found that Index scores predicted desistance from violent reoffending at 1 and 3 years and over the long-term, and desistance from sexual reoffending at 3-years and long-term following discharge.

Other studies have also found that protective factors added incremental predictive validity to risk factors across various outcomes. Lodewijks and colleagues (2008) evaluated the incremental predictive validity of the SAVRY by examining the additive effects of the protective factor subscale on risk factor subscales of the measure. The study found that the protective factor domain added incremental predictive validity for inpatient violence at a correctional treatment among adolescents. Further, Desmarais and colleagues (2012) examined incremental predictive

validity of protective factors using the START, among Canadian, adult forensic patients. The results revealed that total scores on the START Strength Scale added incremental validity to historical items on the HCR-20 for physical aggression. Taken together, these findings suggest the importance of protective factors for predicting forensic patients' behaviour while institutionalized, prior to their release into the community. Moreover, targeting both risk and protective factors may yield even greater results for forensic assessment.

The Nature of the External Subscale of the SAPROF

Upon first glance, some of the findings of the External subscale of the SAPROF appear to be counter-intuitive. Although this subscale demonstrated good internal consistency, it did not correlate with the other subscales of the SAPROF. The External subscale represents external protection; the supervision, structure and treatment needed to manage and reduce risk. Although there appears to be clinical relevance to the External subscale, it was difficult to prove statistically given the findings from the CFA. The first two items (*social network* and *intimate relationship*) are protective factors, whereas the last three items (*professional care, living circumstances,* and *external control*) are related to risk because these items are associated with the level of supervision assigned to the individual based on the overall level of risk posed. One would expect that the first two items and last three items to fall in opposite directions when tested and thus have opposite effects. The first two items did not correlate well with the total External subscale score from the CFA.

Post hoc analyses were conducted to determine whether the External subscale correlated with the other subscales using the first two items (*social network* and *intimate relationship*) and removing the rest. The relationship improved slightly but the External subscale still did not correlate with either Internal or Motivational subscales (r=.05 and r=.25, respectively).

Nevertheless, the External subscale remains a unique predictor in Hypothesis 3, for the prediction of PRN administration over the last year, and upon a six-month follow-up. The use of total scores may be more appropriate for predicting patient behaviour, given the divergence across SAPROF subscales.

The nature of the External subscale has been previously addressed by de Vries Robbe and colleagues (2011). In this study, higher scores on items 15 to 17 represented the presence of greater external control needed to manage the individual's risk and associated with lower levels of protective factors. Conversely, individuals with lower scores on items 15 to 17 likely possess higher levels of protective factors because less external control is required to manage their risk (de Vries Robbe et al., 2011). Further, de Vries Robbe and colleagues (2011), divided SAPROF subscales by static items (items 1 and 2), dynamic improving items (items 3-14), and dynamic decreasing items (items 15-17), rather than by individual subscale for their analyses, in order to conceptualize the direction of change of SAPROF items.

From a clinical perspective, items 15-17 are vital when patients are first hospitalized and other protective factors may still be underdeveloped (de Vries Robbe et al., 2011). Scores on these items are expected to decrease as the patient stabilizes in the hospital and receive treatment. As symptoms stabilize and treatment progresses, items 3-14 develop and items 15-17 are no longer as crucial. Higher scores on Internal and Motivational subscale items reflect a greater balance in internal and social functioning, as such there is a reported shift from extrinsic to intrinsic motivation and control (de Vries Robbe et al., 2011). Thus, there are both theoretical and clinical justifications for the unexpected effects of the External subscale.

The disposition type of the patient may also impact the function of the External subscale. For instance, inpatients will generally score higher on certain items (e.g., *professional care*, *living*

circumstances, and *external control* items), while outpatients will generally have lower scores on these items. Therefore, analyses only including inpatients consisted of a smaller sample size and less variability in External subscale scores, likely making it difficult to detect an effect.

Study Limitations

The current study is not without limitations. Fifty forensic inpatients and outpatients were included, which may be considered small in comparison to previous validation studies of the SAPROF. The methodology of past studies used archival data, whereas the current study used semi-structured interviews involving direct contact with patients. The latter approach is more ecologically valid, as the method and setting which the SAPROF was administered and scored was clinically relevant (Penney et al., 2014). As a result, this required more time and effort to score the SAPROF. To address the concern of having a small sample size, various recruitment strategies were implemented and a more conservative alpha criterion was used in the data analysis.

Some other challenges encountered included difficulty obtaining informed consent due to questionable cognitive capacity and limited ability to concentrate during interviews, given patients' psychiatric symptoms. Further, patients who agreed to participate may differ from those who declined to participate in the study. Higher functioning patients may be more likely to take part in the study, potentially causing the level of protective factors among participants to be greater than that of the general forensic population. However, this is difficult to substantiate, given that protective factors could not be scored for patients who declined to participate.

Another potential limitation of the study is the variability between raters coding risk and protective factor measures. A combination of clinical forensic psychologists and psychology graduate students in training were responsible for scoring the HCR-20 V3, LS/CMI, and PCL-R.

Using different raters, as naturally occurring in a hospital setting, may have reduced the reliability across measures. Still, there may have been greater standardization across HCR-20 V3 ratings because all clinical forensic psychologists at Ontario Shores were required to attend a workshop in 2014 for the implementation of Version 3. Therefore, more consistency between raters may be expected for the HCR-20 V3 compared to other risk measures. Also, the SAPROF was rated by a researcher, whereas clinicians rated risk measures, which may have further influenced the reliability of measures.

An additional drawback of the study was that risk ratings likely had an effect on the level of security assigned to the patient in the upcoming year. Since HCR-20 V3, as well as other risk instruments, are incorporated into each patient's hospital report and shared with the ORB, there is a possibility that risk ratings influenced the ORB's decision about the appropriate level of security. A recent study found that review boards are increasingly considering SPJ evidence such as HCR-20 ratings (Hilton et al., 2016). In contrast, SAPROF scores were not shared with the ORB. This may explain why the HCR-20 V3 predicted changes in privilege and security level but the SAPROF did not. Another consideration is that the ORB is made aware of patients' behaviour during the previous year at the hearing, thus, proxies of recidivism (e.g., PRN administration, institutional misconduct and disposition breaches) may have been predictors of risk management decisions, in addition to outcome variables.

There were some methodological challenges encountered in the present study. Some interested outpatients were unable to commute to the hospital for the research interview. Given this obstacle, three participants opted to partake in a phone interview, while the remaining patients participated in person. As a result, it was difficult to score some items of the SAPS/SANS for participants who completed the phone interview, particularly items that required assessing

nonverbal behaviour. As a result, telephone interviews were sufficient, but not ideal. Post hoc analyses did not find any difference in Composite SAPS/SANS scores between individuals who participated in the in-person interview and telephone interview.

Another procedural issue was the lack of information available for patients who previously resided in a maximum security facility throughout the previous year. This was also an issue for participants who were transferred to a maximum security facility following their ORB hearing. Ontario Shores only houses patients with general and secure dispositions (minimum and medium security), proving difficult to obtain information from other institutions for some study participants.

Lastly, the short-term follow-up of the study may be considered another limitation. Proxies or indicators of recidivism were examined six-months following the interview. It is possible that a six-month period was insufficient to observe changes in low-frequency behaviours, including changes in the PRN medication, institutional misconduct, and disposition breaches. Thus, it is possible that true incident rates of proxies of recidivism were underreported. Other studies have used 1, 2 and 3-year follow-ups; however, the outcomes investigated in these studies were primarily recidivism, which tend to require longer follow-up times (de Vries Robbe et al., 2015a; de Vries Robbe et al., 2015b). Nevertheless, future research exploring the psychometric properties of the SAPROF may consider implementing more extended follow-up periods.

Clinical Implications

Findings from the current study may inform risk assessment, treatment planning, intervention, and risk management decisions and practices implemented by the ORB and clinical practitioners. The inclusion of protective factors in clinical services ensures that the needs of the patient are met, which ultimately results in increased public safety. Protective factor measures are

considered a ‘new frontier’ in violence assessment and steadily gaining momentum in forensic practice (de Ruiter & Nicholls, 2011).

The SAPROF corresponds to GLM principles, which specifies that values and strengths should be the central focus of clinical services (Ward & Brown, 2004; Ward et al., 2012). Unfortunately, this framework has not been extensively studied or implemented in forensic and correctional settings, compared to the RNR model. Although the RNR model is empirically supported, a major drawback is that it focusses on risks and deficits of the individual, uses avoidance based strategies for intervention, and fails to address the function behind offending behaviour (Ward, 2002b). According to the GLM approach, identifying an individual’s values is essential for understanding the function and motivations of offending behaviour (e.g., primary goods, secondary means).

Risk assessments have historically focused on risk factors and deficits, while patient strengths and protective factors have been partially addressed or overlooked altogether (Rogers, 2000). Bolstering strengths rather than repairing weaknesses is a primary role of the clinician (Seligman, 2002), and attending to protective factors is congruent with this role. The same standards of assessment and treatment for the general clinical population must apply to forensic patients. Less reliance on the RNR framework and greater integration of GLM principles and protective factors measures in correctional and forensic settings will ensure that clinical services for NCRMD patients are more thorough and comprehensive.

Considering protective factors in assessment and treatment creates an atmosphere of understanding the patient’s experience, which may result in increased patient engagement and motivation towards recovery. In the current study, patients were asked to recount positive aspects of their lives and what they were currently doing to facilitate their recovery. Patients seemed

appreciative that their strengths were considered, given the tendency for clinical services to focus on identifying and reducing risk. As such, all aspects of the individual were considered using a holistic approach, rather than reductionistic approach.

Psychiatric hospitals such as Ontario Shores, are moving towards a recovery-oriented model of care for forensic patients. This model encapsulates many themes and practices that foster the vision of a 'life worth living'. Examples of this approach include promoting hope, autonomy, meaningful engagement, focussing on strengths, holistic and personalized care, community participation and citizenship, and managing risks by taking calculated risks (McKenna, Furness, Dhital, Park, & Connally, 2014). An emphasis on protective factors can also result in reduced stigmatization and pessimism among clinicians (Rogers, 2000). A common challenge for clinicians is to encourage empowerment, hope, and wellness in conjunction with the limitations imposed by mental health needs and conditions of the legal system (McKenna et al., 2014). The inclusion of the SAPROF in forensic risk assessment is one way in which clinical practice at psychiatric hospitals can operate within the recovery-oriented philosophy.

Assessing risk and developing risk formulations is a primary role of forensic psychologists. Without protective factors, forensic risk assessments are skewed towards patients' deficits. The SAPROF is a useful addition to the routine, risk assessment batteries. In the present study, the SAPRPOF was found to aid in the prediction of disposition breaches. Without this information, decisions may be more conservative, limit individual freedoms, and detain the patient longer than necessary. Inaccurate predictions may lead to prematurely releasing patients into the community, placing the public at risk (Crocker et al., 2010). Most likely, the inclusion of a protective factor instrument in risk assessment may reduce the overclassification of risk level and recidivism rates, having direct implications for risk management decisions made by the ORB.

This is an important implication, given that the role of the review board is to assign the least onerous and restrictive penalty.

Risk assessment should offer more than a prediction of future violent, sexual, or general risk, it should also inform immediate risk management practices (Heilbrun, 1997). Both the ORB and clinical treatment team are involved in making decisions about the most appropriate level of supervision necessary to balance civil rights and public safety. The ORB is the primary authority for making decisions about the appropriate level of security, which is based on information presented in the patient's hospital report (e.g., history, current behaviour, psychiatric symptoms, risk scores), among other sources of information. Recommendations made by the clinical treatment team about the patient's level of security to the ORB were based on information about inappropriate patient behaviour (e.g., critical incidents) over the past year (Martin & Martin, 2016). Also, psychiatrists are increasingly relying on risk instruments when making recommendations to the ORB. Thus, including the SAPROF in risk assessment batteries and hospital reports can aid ORB decision making, since it was shown to predict information influencing risk management recommendations.

The use of protective factor instruments can also guide treatment planning and intervention (Moore & Drennan, 2013). The SAPROF allows clinicians to indicate the particular items that contribute most to the level of protection (e.g., key factors on the SAPROF), as well as worthy targets for intervention (e.g., goal factors on the SAPROF). For instance, compliance and insight into medication may be key to reduce risk of reoffending, whereas strengthening self-control skills may be a goal for intervention (e.g., receiving treatment for emotional dysregulation). Additionally, items on the External subscale can be useful for guiding risk management strategies and decision making regarding changes in security levels especially for

inpatients, given that these items identify environmental factors necessary to protect both the patient and the public (de Vries Robbe, et al., 2011). Identifying these factors can be useful for tailoring treatment to appropriate needs, since each patient possesses their unique offending profile of risk and protective factors.

SAPROF scores can also inform the treatment team of the most appropriate type of intervention. According to the GLM, treatment should not be limited to targeting criminogenic needs (i.e. factors directly related to offending behaviour). The purpose of treatment should also serve to bolster strengths in order to help the individual lead a meaningful and fulfilling life (Ward & Brown, 2004). Given that both risk and protective factors were shown to predict disposition breaches, decreasing an individual's risk may not be sufficient to reduce this problematic behaviour. Therefore, in addition to reducing risk factors, protective factors will likely need to be augmented. This suggests that increasing prosocial behaviour may be an important consideration for decreasing antisocial behaviour.

Future Directions

There are multiple possibilities for future research directions related to the SAPROF. The current study relied on semi-structured interviews and file information to score the SAPROF. Future studies should consider more comprehensive methods of gathering information, such as in-person interviews with patients and incorporating collateral information from patients' family members and treatment team, rather than solely relying on file reviews. Implementing these methodologies may serve to improve the accuracy of SAPROF scores. Further, the present study examined several outcome variables over a six-month follow-up, which was not previously explored in the literature. Investigating these outcomes over a longer period may be worthwhile to

establish whether changes in the outcome variables (e.g., PRN medication, institutional misconduct, and disposition breaches) are maintained long-term.

The current study relied on total scores, rather than ratings/judgements (e.g., high, moderate, low risk/protection) for the majority of statistical analyses. The rationale for using total scores was that they are a more objective measure, since it requires less clinician bias than assigning ratings. Also, total scores allow for more variability between scores and may detect nuances that ratings cannot. It is recommended to use risk ratings/judgements for SPJ tools such as the SAPROF and HCR-20 V3, rather than total scores (Logan & Johnstone, 2012). Nevertheless, a large meta-analysis found that total scores of SPJ tools had greater predictive validity than risk ratings (Hanson & Morton-Bourgon, 2009). This may be due to the greater variability in total scores compared to risk ratings which are distinct categories. Therefore, the clinical utility of total scores may be a future research consideration. Also, the current study calculated and reported an Index Score, which is the corrected risk score computed by subtracting SAPROF total scores from HCR-20 V3 total scores. Index Scores revealed good predictive validity for PRN administrations and disposition breaches upon the six-month follow-up, and may be worthy of further exploration.

A wider issue worthy of future research involves identifying the complex interplay between risk and protective factors in relation to offending behaviour and patient behaviour, in general. As noted earlier, there is a debate in the literature about the role and function of protective factors. Different models have been proposed, including considering protective factors as mediators, moderators (having indirect effects), and main effects (having direct effects) relative to risk factors (de Vries Robbe, 2014; Fitzpatrick, 1997). The conceptualization of protective factors in relation to risk factors remains unclear, and an agreed-upon definition is yet to be

determined. The findings from the current study support an additive or incremental model of protection, suggesting that the relationship between protection and risk is additive or cumulative. Thus, the presence of protective factors may serve to buffer the effect of risk on a negative or maladaptive outcome (e.g., recidivism, institutional misconduct, disposition breaches). Continuing to delineate the dynamic relation between protection, risk, and offending behaviour may elucidate the reasons for desistance from offending behaviour.

Future research may also consider individual protective factors that serve to protect NCRMD patients, rather than the quantity of protective factors. Analyses for the present study mainly focused on the sum of protective factors (e.g., SAPROF total scores), as predictive of several outcomes. It is possible that particular individual protective factors accounted for the association between SAPROF scores and the outcomes of interest (Zeng et al., 2015). For instance, a recent meta-analysis demonstrated that a higher level of intelligence is predictive of lower levels of offending among high-risk and low-risk groups (Ttofi et al., 2016). Additionally, Coid and colleagues (2015), de Vries Robbe and colleagues (2011), and Yoon and colleagues (2011) all found that self-control accounted for desistence from violence among forensic patients and sexual offenders. Similarly, Davoren and colleagues (2013) found that self-control predicted which patients received a conditional discharge. Therefore, it is possible that a single protective factor such as intelligence or self-control may protect against offending behaviour and predict successful progression through the mental health system, and may be worthy of further investigation.

Lastly, the present study was a prospective and retrospective cross-validation study, and exploratory in nature. As with all exploratory research, the current study requires replication to ensure its generalizability to other populations, settings and across outcome variables.

Conclusion

The study found some evidence for various forms of validity and reliability and therefore makes an important contribution to the literature. The main finding was that the SAPROF approached significance for incrementally predicting disposition breaches over the past year and institutional misconducts over the six-month follow-up, when added to the HCR-20 V3. As expected, higher scores on the HCR-20 V3 and lower scores on the SAPROF were indicative of problematic patient behaviour. Given that the SAPROF increased the accuracy of violence risk assessment, its addition to conventional risk assessment batteries administered by clinical forensic psychologists is warranted. Further, the inclusion of SAPROF scores in hospital reports can inform ORB decisions about the most appropriate level of supervision necessary to manage risk. Incorporating protective factors into clinical services is in line with a recovery-oriented model of care. This model of care promotes patient strengths, empowerment and hope, rather than defining forensic patients by risk factors and deficits.

Appendix A: Consent Form

Name of Study: The Reliability and Validity of the SAPROF among Forensic Mental Health Patients

Investigators: Dr. Lisa Marshall (principal investigator)
Sandy Oziel (MA)
Dr. David Day

You have been invited to participate in this study. Before you agree to participate, it is important that you read the following information and ask as many questions as necessary to be sure that you understand the study.

Purpose:

- We are interested in understanding how patient strengths influence patient behaviour, psychiatric medication and Ontario Review Board (ORB) decisions.
- Including patient strengths in risk assessments may provide a balanced account of patient progress and help staff better manage patient risks.
- The study will take place in the Forensic Program at Ontario Shores Centre for Mental Health Sciences.
- Our aim is to have 100 patients participating in the study over an 8 month period.
- The current study is a joint project between Ontario Shores and Ryerson University. The study will be led by Sandy Oziel, a graduate student supervised by Drs. Marshall and Day. The results will contribute to a dissertation for her PhD degree.

Involvement: If you agree to participate in this study, you agree to an interview with Sandy and allow for your patient file (i.e., data that is collected during your routine care at Ontario Shores) to be reviewed for research purposes. The research interview and file information collected will cover areas related to mental health, current treatments, participation in vocational activities, social supports, daily functioning and your current legal status under the ORB. The research interview will also include a brief evaluation of intelligence. Participation will last approximately 90 minutes in total and will take place over 1-2 sessions.

Eligibility: You are eligible to participate in this study if you are a patient in the Forensic Program at Ontario Shores, found not criminally responsible. Patients who require a substitute decision maker for treatment only are eligible to participate.

Risks: There are no known physical or psychological risks that results from this study. However, the personal nature of some questions asked may cause you to feel uncomfortable or upset. You may skip a question or stop participation at any time, either temporarily or permanently.

Benefits: You will not directly benefit from this research. However, the knowledge that you provide may help us to better understand patient strengths that relate to problematic behaviours and risk management decisions made by the ORB. This knowledge may help to improve existing risk assessments and treatment to better serve individuals in the forensic mental health system.

Incentive: You will receive a $10.00 gift card as incentive for your participation.

Voluntary Participation: Your participation in this study is voluntary. You may choose to withdraw from the study at any time. In addition, the researcher may end your participation at any time, at their discretion. Your choice not to participate or to withdraw will not affect any treatment services that you might be receiving at Ontario Shores now or in the future. If you wish to withdraw from the study at a later date, please contact Dr. Lisa Marshall. If you choose to withdraw, all study information relating to you will be destroyed and not used in the study. If you decide to stop participating at any point, you will receive partial payment based on the amount of participation.

Confidentiality: All information gathered from you will be kept confidential to the full extent allowed by law. Guidelines set out by law and/or the College of Psychologists of Ontario states that confidentiality will not be kept under the following circumstances:

- The participant is at immediate risk of harming him/herself or others.
- There is a reasonable belief of emotional neglect, physical neglect and/or sexual abuse of a minor.
- The participant reports having been sexually abused by a regulated health professional and provides the name of that professional.

As part of a continuing review of research at Ontario Shores, your study records may be reviewed by the Ontario Shore's Research Ethics Board. A person from the research ethics team may contact you (if your contact information is available) to ask you questions about the research study and consent to participate. The person reviewing your file must maintain your confidentiality allowed by law.

The information you provide will be stored securely and confidentially (except as required by law) in encoded computer files and in a locked cabinet (hardcopy forms) at Ontario Shores with restricted access; only the principal investigator, co-investigators and supervised research assistant will have access to the information. Your name or any other personal identifying information on files and test forms will be changed to a numerical code and will not be used in any presentations, reports or publications from this study. The information will be stored for 7 years, after which it will shredded and destroyed.

Additional Information: The results of the study will be presented during community meetings at Ontario Shores. If you would like to receive a summary of the results, please notify Sandy, and the results will be mailed to Ontario Shores and distributed to you.

An audio recording of the research interview will be taken for research purposes to score the assessment tools. The audio recording is optional and your decision not to be recorded does not prevent you from participating in the study. The audio recording will be stored at Ontario Shores, reviewed only by researchers associated with the study and destroyed once all information collected for the study is complete.

If you have any questions regarding your rights, you may contact Dr. Ron Heslegrave, Chair, Research Ethics Board, Ontario Shores.

CONSENT TO PARTICIPATE

I, _____, have read (or had read to me) the consent form for the study entitled 'The Reliability and Validity of the SAPROF among Forensic Mental Health Patients.' My role as a study participant is to help the investigators collect information on patient strengths. My questions, if any, have been answered to my satisfaction. By signing this form I do not waive any of my rights.

Dr. Ron Heslegrave, Chair, Research Ethics Board, Ontario Shores Centre for Mental Health Sciences, may be contacted by research subjects to discuss patient rights.

Also, if you have questions regarding your rights as a human subject and participant in this study, you may contact the Ryerson University Research Ethics Board for information.

I consent to (check all that apply):

☐ Participate in the research interview
☐ Research staff accessing my health records
☐ Audio recording of research interview

I hereby grant permission to Sandy Oziel to audio record the research interview, for research purposes only.

_____
(Name of participant)

I have been offered a signed copy of this form.

Research Volunteer:	**Person Obtaining Consent:**
_____	_____
(Print name)	(Print name)
_____	_____
(Signature)	(Signature)
_____	_____
(Date)	(Date)

Appendix B: Ontario Review Board Condition Levels

Hospital and Grounds Conditions

Level 1 - Accompanied (please enter in the name of the person accompanying the patient) *M = mandatory for medical, legal and compassionate uses.

Level 2 - Indirectly Supervised: Up to 1/2 hour, 0800 - 1600 hrs, contact at 15 minute intervals

Level 3 - Indirectly Supervised: Up to 1 hour, 0800 - 1800 hrs, contact at 30 minute intervals

Level 4 - Indirectly Supervised: Up to 2 hours, 0800 - 2100, contacts at 60 minute intervals

Level 5 - Indirectly Supervised: Up to 4 hours, 0800 - 2100, contacts at 2 hour intervals

Level 6 - Indirectly Supervised: Up to 8 hours, 0800 - 2100, contacts at 4 hours intervals

Level 7 - Indirectly Supervised: Up to 12 hours, 0800 - 2100, contacts at 6 hour intervals

Community Conditions

Level 8 - Accompanied (please enter the name of the person accompanying the patient) *M = mandatory for medical, legal and compassionate uses.

Level 9 - Indirectly Supervised: Up to 3 hours, 0800 - 2100, contacts at 1 hour intervals

Level 10 - Indirectly Supervised: Up to 6 hours, 0800 - 2100, contacts at 2 hour intervals

Level 11 - Indirectly Supervised: Up to 8 hours, 0800 - 2100, contacts at 4 hour intervals

Level 12 - Indirectly Supervised: Up to 12 hours, 0800 - 2100, contacts at 6 hour intervals

Level 13 - Indirectly Supervised: Up to one (1) week, contacts at 12 hour intervals daily

Level 14 - Community Residence - Detention Order

References

Aas, M. (2010). Review Global Assessment of Functioning (GAF): Properties and frontier of current knowledge. *Annals of General Psychiatry*, *9*, 1-11

Abidin, Z., Davoren, M., Naughton, L., Gibbons, O., Nulty, A., & Kennedy, H. (2013). Susceptibility (risk and protective) factors for in-patient violence and self-harm: Prospective study of structured professional judgement instruments START and SAPROF, DUNDRUM-3 and DUNDRUM-4 in forensic mental health services. *BioMed Central Psychiatry*, *13*, 1-18.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Andreasen, N., (1983). *The Scale for the Assessment of Negative Symptoms (SANS)*. University of Iowa, Iowa City.

Andreasen, N. (1984). *The Scale for the Assessment of Positive Symptoms (SAPS).* University of Iowa, Iowa City.

Andrews, D. A., & Bonta, J. (2000). *The level of service inventory-revised*. Toronto, Canada: Multi-Health Systems.

Andrews, D., & Bonta, J. (2006). *The psychology of criminal conduct (4th ed.)*. Newark, NJ: LexisNexis/Matthew Bender.

Andrews, D., Bonta, J., & Hoge, R. (1990). Classification for effective rehabilitation. *American Association for Correctional Psychology*, *17*, 19-52.

Andrews, D. A., Bonta, J., & Wormith, S. J. (2000). *Level of service/case management inventory: LS/CMI*. Multi-Health Systems.

Andrews, D., Bonta, J., & Wormith, S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, *52*, 7-27. doi: 10.1177/0011128705281756

Andrews, D., Bonta, J., & Wormith, J. (2011). The Risk-Need-Responsivity (RNR) model, does adding the Good Lives Model contribute to effective crime prevention? *Criminal Justice and Behavior*, *38*, 735-755.

Andrews, D., & Dowden, C. (2006). Risk principle of case classification in correctional treatment, a meta-analytic investigation. *Offender Therapy and Comparative Criminology*, *50*, 88-100. doi: 10.1177/0306624X05282556

Arbuckle, J. (2009). *Amos 18 user's guide.* Chicago, IL: SPSS Inc.

Attrill, G., & Liell, G. (2007). Offenders views on risk assessment. In N. Padfield (Ed.) *Who to release? Parole, fairness and criminal justice* (pp. 191-201). Cullompton, UK: Willan.

Barbaree, H., Seto, M., Langton, C., & Peacock, E. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, *28*, 490-521.

Bradley, B., Sepehri, A., Hecht, F., VandenBroek, A., Ryan, J., & Drabman, R. (2001). A confirmatory factor analysis of the WISC-III in a clinical sample with cross-validation in the standardization sample. *Child Neuropsychology*, *7*, 104-116.

Barnao, M. (2013). The Good Lives Model tool kit for mentally disordered offenders. *Journal of Forensic Practice*, *15*, 157-170. doi: 10.1108/JFP-07-2012-0001

Barnao, M., Robertson, P., & Ward, T. (2010). Good Lives Model applied to a forensic population. *Psychiatry, Psychology and Law*, *17*, 202-217.

Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Bentler, P., & Bonett, D. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588-606.

Bill C-30. (1992). An Act to amend the Criminal Code and to amend the National Defense Act and the Young Offenders Act in consequence thereof (assented to 1991, c. 43, s. 4, proclaimed in force February 4, 1992).

Bloom, H., & Schneider, R. (2006). *Mental disorder and the law: A primer for legal and mental health professionals*. Toronto, Ontario: Irwin Law.

Boer, B. & Blacker, J. (2010). The importance of ecological validity for risk assessment. In R. Hicks (Ed.), *Personality and Individual Differences: Current Directions* (pp. 271-282). Bowen Hills, Australia: Australian Academic Press.

Boer, D., Hart, S., Kropp, P., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk - 20: Professional guidelines for assessing risk of sexual violence.* Vancouver: The Mental Health, Law, & Policy Institute

Bonta, J., & Andrews, D. (2007). *Risk-Need-Responsivity model for offender assessment and rehabilitation 2007-06*. Her Majesty the Queen in Right of Canada.

Bonta, J., Wallace-Capretta, S., & Rooney, J. (2000). A quasi-experimental evaluation of an intensive rehabilitation supervision program. *Criminal Justice and Behavior*, *27*, 312-329.

Borum, R., Bartel, P., Forth, A. (2003). *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY)* version 1.1. Florida: University of South Florida.

Bouch, J. & Marshall, J. (2003). S-RAMM: Suicide Risk Assessment and Management Manual (Research Edition). *Glamorgan, Wales: The Cognitive Centre Foundation*.

Bouman, Y., de Ruiter, C., & Schene, A. (2008). Quality of life of violent and sexual offenders in community-based forensic psychiatric treatment. *The Journal of Forensic Psychiatry & Psychology*, *19*, 484-501.

Bowerman, B., & O'Connell, R. (1990). *Linear statistical models: An applied approach*. Boston: PWS-Kent.

Braithwaite, E., Charette, Y., Crocker, A., & Reyes, A. (2010). The predictive validity of clinical ratings of the Short-Term Assessment of Risk and Treatability (START). International *Journal of Forensic Mental Health*, *9*, 271-281. doi: 10.1080/14999013.2010.534378.

Brown, T. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions*, *154*, 136-162.

Campbell, M., French, S., Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, *36*, 567-590.

Canadian Bar Association. (2013). *Bill C-54: Not Criminally Responsible Reform Act*. Ottawa, ON.

Canadian Psychiatric Association. (2013). The consequences of Bill C-54, the Not Criminally Responsible Reform Act. *Immediate Information Release.* Ottawa, ON.

Chu, C., Thomas, J., Ogloff, J., & Daffern, M. (2011). The predictive validity of the Short-Term Assessment of Risk and Treatability (START) in a secure forensic hospital: Risk factors and strengths. *International Journal of Forensic Mental Health*, *10*, 337-345. doi:

10.1080/14999013.2011.629715

Cleves, M. (2002). Comparative assessment of three common algorithms for estimating the variance of the area under the nonparametric receiver operating characteristic curve. *The Stata Journal*, *3*, 280-289.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*, 155-159.

Coid, J., Kallis, C., Doyle, M., Shaw, J., & Ullrich, S. (2015). Identifying causal risk factors for violence among discharged patients. *PloS One*, *10*, 1-17. doi:10.1371/journal.pone.0142493

Corcoran, K., & Fischer, J. (1987). *Measures for clinical practice: A sourcebook*. USA: The Free Press, A Division of Macmillan, Inc.

Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*, 98-104.

Costa, F., Jessor, R., & Turbin, M. (1999). Transition into adolescent problem drinking: The role of psychosocial risk and protective factors. *Journal of Studies on Alcohol*, *60*, 480-490.

Craig, L., Browne, K., Stringer, I., & Beech, A. (2005). Sexual recidivism: A review of static, dynamic and actuarial predictors. *Journal of Sexual Aggression*, *11*, 65-84.

Crocker, A. Braithwaite, E., Cote, G., Nicholls, T., & Seto, M. (2011). To detain or to release? Correlates of dispositions for individuals declared Not Criminally Responsible on Account of Mental Disorder. *Canadian Journal of Psychiatry*, *56*, 293-302.

Crocker, A., Charette, Y., Seto, M., Nicholls, T., Cote, G., & Caulet, M. (2015). The National Trajectory Project of individuals found Not Criminally Responsible on Account of Mental Disorder in Canada. Part 3: Trajectories and outcomes through the forensic system. *Canadian Journal of Psychiatry*, *60*, 117-126.

Crocker, A., Nicholls, T., Cote, G., Latimer, E., & Seto, M. (2010). Individuals found Not Criminally Responsible on account of Mental Disorder: Are we providing equal protection and equivalent access to mental health services across Canada? *Canadian Journal of Community Mental Health, 29,* 47-54.

Crocker, A., Nicholls, T., Seto, M., Charette, Y., Cote, G., & Caulet, M. (2015a). The National Trajectory Project of individuals found Not Criminally Responsible on Account of Mental Disorder in Canada. Part 2: The people behind the label. *Canadian Journal of Psychiatry, 60,* 106-116.

Crocker, A., Nicholls, T., Seto, M., Cote, G., Charette, Y., & Caulet, M. (2015b). The National Trajectory Project of individuals found Not Criminally Responsible on Account of Mental Disorder in Canada. Part 1: Context and Methods. *Canadian Journal of Psychiatry, 60,* 98-105.

Davoren, M., Abidin, Z., Naughton, L., Gibbons, O., Nulty A., Wright, B., & Kennedy, H. (2013). Prospective study of factors influencing conditional discharge from a forensic hospital: The DUNDRUM-3 programme completion and DUNDRUM-4 recovery structured professional judgment instruments and risk. *BMC Psychiatry, 13,* 185-201.

Day, D., Wanklyn, S., & Yessine, A. (2013). A review of terminological, conceptual, methodological issues in the developmental risk factor literature for antisocial and delinquent behavior. *Child Youth Care Forum, 43,* 97-112. doi: 10.1007/s10566-013-9227-9.

DeLong, E., DeLong D., Clarke-Pearson, D. (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44,* 837-845.

de Ruiters, C., & Nicholls, T. (2011). Protective factors in forensic mental health: A new frontier. *International Journal of Forensic Mental Health, 10*, 160-170. doi:10.1080/14999013.2011.600602

de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbe, M. (2009). SAPROF. Guidelines for the assessment of protective factors for violence risk. Utrecht: Forum Educatief.

de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbe, M. (2009). Guidelines for the assessment of protective factors for violence risk. English version. Utrecht, The Netherlands: Forum Educatief.

de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbe, M. (2012). SAPROF. Guidelines for the assessment of protective factors for violence risk (2nd ed.). Utrecht: Van der Hoeven Stichting.

de Vogel, V., de Vries Robbe, M., de Ruiter, C., & Bouman, Y. (2011). Assessing protective factors in forensic psychiatric practice: Introducing the SAPROF. *International Journal of Forensic Mental Health*, *10*, 171-177. doi: 10.1080/14999013.2011.600230

de Vries Robbe, M. (2014). *Protective factors, validation of the Structured Assessment of Protective Factors for Violence Risk in forensic psychiatry* (Unpublished doctoral dissertation). Radboud University Nijmegen, Netherlands.

de Vries Robbe, M., & de Vogel, V. (2013). Protective factors for violence risk: Bringing balance to risk assessment. In C. Logan, & L. Johnstone (Eds.), *Managing Clinical Risk: A guide to effective practice* (pp. 293-310). London: Routledge.

de Vries Robbe, M., & de Vogel, V. (2014). *SAPROF Interview Self-Appraisal* (SAPROF-ISA). Utrecht: Van der Hoeven Kliniek.

de Vries Robbe, M., de Vogel, V., & de Spa, E. (2011). Protective factors for violence risk in forensic psychiatric patients: A retrospective validation study of the SAPROF. *International Journal of Forensic Mental Health*, *10*, 178-186. doi: 10.1080/14999013.2011.600232

de Vries Robbe, M., de Vogel, V., & Douglas, K. (2013). Risk factors and protective factors: A two-sided dynamic approach to violence risk. *The Journal of Forensic Psychiatry & Psychology*, *24*, 440-457.

de Vries Robbe, M., de Vogel, V., Douglas, K., & Nijman, H. (2015a). Changes in dynamic risk and protective factors for violence during inpatient forensic psychiatric treatment: Predicting reductions in post-discharge community recidivism. *Law and Human Behavior, 39*, 53-61. doi:10.1037/lhb0000089

de Vries Robbe, M., de Vogel, V., Koster, K., & Bogaerts, S. (2015b). Assessing protective factors for sexually violent offending with the SAPROF. *Sexual Abuse: A Journal of Research and Treatment*, *27*, 51-70.

de Vries Robbe, M., de Vogel, V., Wever, E., Douglas, K., & Nijman, H. (2016). Risk and protective factors for inpatient aggression. *Criminal Justice and Behavior*, 1-22, doi: 10.1177/0093854816637889

Desmarais, S., Hucker, S., Brink, J., & De Freitas, K. (2008). A Canadian example of insanity defence reform: Accused found Not Criminally Responsible before and after the Winko decision. *International Journal of Forensic Mental Health, 7*, 1-14.

Douglas, K., Blanchard, A., Guy, L., Reeves, K., & Weir, J. (2010). HCR-20 Violence risk assessment scheme: Overview and annotated bibliography (online).

Douglas, K., Guy, L., & Hart, S. (2009). Psychosis as a risk factor for violence to others: A meta-analysis. *Psychological Bulletin*, *135*, 679-706.

Douglas, K., Hart, S., Webster, C., & Belfrage, H. (2013). *HCR-20 V3: Assessing risk of violence – User guide*. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Douglas, K., & Kropp, P. (2002). A prevention-based paradigm for violence risk assessment. Clinical and research applications. *Criminal Justice and Behavior, 29*, 617-658.

Douglas, K., Ogloff, J., Nicholls, T., & 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., & Reeves, K. (2010). Historical-Clinical-Risk Management-20 (HCR-20) Violence risk assessment scheme: Rationale, application, and empirical overview. In R.K. Otto, & K.S. Douglas (Eds.), *Handbook of violence risk assessment. International Perspectives on forensic mental health* (pp 147-185). New York: Routledge.

Douglas, K., & Skeem, J. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law, 11*, 347-383.

Douglas, K., & Webster, C. (1999). The HCR-20 violence risk assessment scheme concurrent validity in a sample of incarcerated offenders. *Criminal Justice and Behavior, 26*, 3-19.

Douglas, K. S., Yeomans, M., & Boer, D. P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior, 32*, 479-510.

Doyle, M., Coid, J., Archer-Power, L., Dewa, L., Hunter-Didrichsen, A., Stevenson, R.,…Shaw, J. (2014). Discharges to prison from medium secure psychiatric units in England and Wales. *The British Journal of Psychiatry, 205*, 177-183. doi: 10.1192/bjp.bp.113.1366222

Dunlap, W. (1999). A program to compute McGraw and Wong's common language effect size

indicator. *Behavior Research Methods, Instruments, & Computers, 31*, 706-709.

Elbogen, E., & Johnson, S. (2009). The intricate link between violence and mental disorder: Results from the National Epidemiologic Survey on alcohol and related conditions. *Archives of general Psychiatry, 66*, 152-161.

Farrington, D. (2000). Explaining and preventing crime: The globalization of knowledge-the American Society of Criminology 1999 Presidential Address. *Criminology, 38,* 1-24. doi: 10.1111/j.1745-9125.2000.tb00881.x

Farrington, D. & Loeber, R. (2000). Epidemiology of juvenile violence. *Child and Adolescent Psychiatric Clinics of North America, 9,* 733-748.

Field, A. (2009). *Discovering statistics using SPSS*. Sage publications: London.

Fitzpatrick, K. (1997). Fighting among America's youth: A risk and protective factors approach. *Journal of Health and Social Behavior, 38*, 131-148.

Flynn, G., O'Neill, C., McInerney, C., & Kennedy, H. (2011). The DUNDRUM-1 structured professional judgment for triage to appropriate levels of therapeutic security: Retrospective-cohort validation study. *BMC psychiatry, 11,* 1-11. doi:10.1186/1471-244X-11-43

Gammelgård, M., Koivisto, A., Eronen, M., & Kaltiala-Heino, R. (2008). The predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) among institutionalised adolescents. *The Journal of Forensic Psychiatry & Psychology, 19*, 352-370.

Gatignon, H. (2003). *Statistical analysis of management data*. Boston, MA: Kluwer Academic Publishers.

Girard, T., Axelrod, B., Patel, R., & Crawford, J. (2015). Weschler Adult Intelligence Scale-IV dyads for estimating global intelligence. *Assessment, 22*, 441-448. doi: 10.1177/1073191114551551

Girard, L., & Wormith, J. S. (2004). The predictive validity of the Level of Service Inventory-Ontario Revision on general and violent recidivism among various offender groups. *Criminal Justice and Behavior, 31*, 150-181.

Grann, M., Belfrage, H., & Tengstrom, A. (2000). Actuarial Assessment of Risk for Violence Predictive Validity of the VRAG and the Historical Part of the HCR-20. *Criminal Justice and Behavior*, *27*, 97-114.

Gray, N., Benson, R., Craig, R., Davies, H., Fitzgerald, S., & Huckle, P., Maggs, R. (2011). The Short-Term Assessment of Risk and Treatability (START): A prospective study of inpatient behaviour. *International Journal of Forensic Mental Health*, *10*, 305-313. doi: 10.1080/14999013.2011.631692

Green, T., Batson, A., & Gudjonsson, G. (2010). The development and initial validation of a service-user led measure for recovery of mentally disordered offenders. *Journal of Forensic Psychiatry & Psychology*, *22*, 252-265.

Gregoire, J., & Jungers, C. (2007). *The counselor's companion: What every beginning counselor needs to know*. Mahwah, NJ: Routledge.

Griffin, H., Beech, A., Print, B., Bradshaw, H., & Quayle, J. (2008). The development and initial testing of the AIM2 framework to assess risk and strengths in young people who sexually offend. *Journal of Sexual Aggression*, *14*, 211-225.

Griffin, H., & Vettor, S. (2012). Predicting sexual re-offending in a UK sample of adolescents with intellectual disabilities. *Journal of Sexual Aggression, 18*, 64-80.

Guy, L. (2008). *Performance indicators of the structured professional judgement approach for assessing risk for violence to others: A meta-analytic survey.* Unpublished dissertation, Simon Fraser University, Burnaby, British Columbia, Canada.

Hales, H., & Gudjonsson, G. (2004). Effect of ethnic differences on the use of prn (as required) medication on an inner London medium secure unit. *The Journal of Forensic Psychiatry & Psychology*, *15*, 303-313.

Hanby, L. (2013). *A longitudinal study of dynamic risk, protective factors, and criminal recidivism: Change over time and the impact of assessment timing* (Doctoral dissertation, Carleton University Ottawa, Canada).

Hanley, J., & McNeil, B. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, *148*, 839-843.

Hanson, R. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology*, *50*, 172-192. doi: 10.1037/a0015

Hanson, R., & Bussière, M. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, *66*, 348-362.

Hanson, R., & Morton-Bourgon, K. (2004). *Predictors of Sexual Recidivism: An Updated Meta-analysis*. Public Works and Government Services Canada. Cat. No.: PS3-1/2004-2E PDF. ISBN: 0-662-36397-3.

Hanson, K., & Morton-Bourgon, K. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 studies. *Psychological Assessment*, *21*, 1-21.

Hanson, R., & Thornton, D. (2003). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, *24*, 119-136.

Hare, R. (1991). *The Hare Psychopathy Checklist-Revised* (PCL-R). Toronto, ON: Multi-Health Systems.

Hare, R. (2003). *The Hare Psychopathy Checklist—Revised* (PCL-R, 2nd ed.). Toronto: Multi-Health Systems.

Hare, R., Clark, D., Grann, M., & Thornton, D. (2000). Psychopathy and the predictive validity of the PCL-R: An international perspective. *Behavioral Sciences and the Law, 18*, 623-645.

Hare, R., Harpour, T., Hakistan, R., Forth, A., Hart, S., & Newman, J. (1990). The Revised Psychopathy Checklist: Reliability and factor structure. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *2*, 338–341.

Harpur, T., Hare, R., & Hakstian, R. (1989). Two-factor conceptualization of psychopathy: Construct validity and assessment implications. *Psychological Assessment: Journal of Consulting and Clinical Psychology*, *1*, 6-17.

Harris, G., Rice, M., & Quinsey, V. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20*, 315-335.

Hays, J., Reas, D., Shaw, B. (2002). Concurrent validity of the Wechsler Abbreviated Scale of Intelligence and the Kaufman Brief Intelligence Test among psychiatric inpatients. *Psychological Reports, 90*, 355-359.

Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior*, *21*, 347.

Heilbrun, K., Yasuhara, K., & Shah, S. (2010). Violence risk assessment tools: Over- view and critical analysis. In R.K. Otto & K.S. Douglas (Eds.), *Handbook of violence risk assessment. International perspectives on forensic mental health* (pp. 1–17). New York, NY: Routledge.

Hiday, V. (1990). Dangerousness of civil commitment candidates: A six-month follow-up. *Law and Human Behavior, 14*, 551-567.

Hilton, Z., & Simmons, J. (2001). The influence of actuarial risk assessment in clinical judgements and tribunal decisions about mentally disordered offenders in maximum security. *Law and Human Behavior, 25*, 393-408.

Hilton, Z., Simpson, A., & Ham, E. (2016). The increasing influence of risk assessment on forensic patient review board decisions. *Psychological Services*, *13*, 223-231.

Hoge, R., Andrews, D., & Leschied, A. (1996). An investigation of risk and protective factors in a sample of youthful offenders. *Journal of Child Psychology and Psychiatry, 37,* 419-424.

Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural equation modelling: Guidelines for determining model fit. *Journal of Business Research Methods, 6*, 53–60.

Hosmer, D., Lemeshow, S., & Sturdivant, R. (2000). Model-building strategies and methods for logistic regression. *Applied Logistic Regression, Third Edition*, 89-151.

Hoyle R. (1995). *Structural equation modeling: concepts, issues and applications.* Thousand Oaks (CA): Sage.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Kay, S., Fiszbein, A., & Opfer, L. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin, 13*, 261-276.

Kay, S., Opler, L., & Lindenmayer, J. (1988). Reliability and validity of the positive and negative syndrome scale for schizophrenics. *Psychiatry Research, 23*, 99-110.

Kazdin, A. (2003). *Research design in clinical psychology, fourth edition*. Boston, MA: Allyn and Bacon.

Klein, V., Yoon, D., Briken, P., Turner, D., Spehr, A., & Rettenberger, M. (2012). Assessment of accused juvenile sex offenders in Germany: A comparison of five different measures. *Behavioral Sciences and the Law*, *30*, 181-195. doi: 10.1002/bsl.2006

Kline, P. (2000). *The handbook of psychological testing*. London: New York Routledge.

Kline, R. (2010). *Principles and practice of structural equation modeling (3rd ed.).* New York, New York: Guilford Press.

Kroner, D., & Mills, J. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior*, *28*, 471-489.

Kropp, P., Hart, S., Webster, C., & Eaves, D. (1994). *Manual for the Spousal Assault Risk Assessment Guide*. Vancouver, BC: British Columbia Institute on Family Violence.

Landis, J. & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

LeBel, T., Burnett, R., Maurna, S., & Bushway, S. (2008). The 'chicken and egg' of subjective and social factors in desistance form crime. *European Journal of Criminology*, *5*, 131-159.

Leistico, A., Salekin, R., DeCoster, J., & Rogers, R (2008). A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. *Law and Human Behavior*, *32*, 28-45. doi: 10.1007/s10979-007-9096-6

Lindsay, W. R., Ward, T., Morgan, T., & Wilson, I. (2007). Self-regulation of sex offending, future pathways and the Good Lives Model: Applications and problems. *Journal of sexual aggression*, *13*, 37-50.

Livingston, J., Wilson, D., Tien, G., & Bond, L. (2003). A follow-up study of persons found Not Criminally Responsible on Account of Mental Disorder in British Columbia. *The Canadian Journal of Psychiatry*, *48*. 408-415.

Lodewijks, H., Doreleijers, T., 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.

Lodewijks, H., de Ruiter, C., & Doreleijers, T. (2010). The impact of protective factors in desistance from violence offending: A study in three samples of adolescent offenders. *Journal of Interpersonal Violence*, *25*, 568-587.

Logan, C., & Johnstone, L. (2012). *Managing clinical risk: A guide to effective practice* (Vol. 3). Routledge.

Lussier, P., Verdun-Jones, S., Deslauriers-Varin, N., Nicholls, T., & Brink, J. (2010). Chronic violent patients in an inpatient psychiatric hospital: Prevalence, description, and identification. *Criminal Justice and Behaviour*, *37*, 5-28. doi: 10.1177/0093854809347738

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130-149.

MacDonald, Z. (2002). Official crime statistics: Their use and interpretation. *The Economic Journal*, *112*, 85-106.

Martin, K., & Martin, E. (2016). Factors influencing treatment team recommendations to review tribunals for forensic psychiatric patients. *Behavioral Sciences and the Law*, 1-13. doi: 10.1002/bsl.2244

Maruna, S., & LeBel, T. (2003). Welcome home? Examining the "re-entry court" concept from a strength-based perspective. *Western Criminology Review, 4*, 91-107.

McKenna, B., Furness, T., Dhital, D., Park, M., & Connally, F. (2014). Recovery-oriented care in a secure mental health setting: "Striving for a good life." *Journal of Forensic Nursing, 10*, 63-69.

Mezey, G. C., Kavuma, M., Turton, P., Demetriou, A., & Wright, C. (2010). Perceptions, experiences and meanings of recovery in forensic psychiatric patients. *The Journal of Forensic Psychiatry & Psychology*,*21*, 683-696.

Miller, H. (2006). A dynamic assessment of offender risk, needs, and strengths in a sample of pre-release general offenders. *Behavioral Sciences and the Law*, *24*, 767-782. doi: 10.1002/bsl.728.

Mills, J., Kroner, D., & Morgan, R. (2011). *The clinician's guide to violence risk assessment.* New York, NY: Guilford Press.

Moore, E., & Drennan, G. (2013). Complex forensic case formulation in recovery-oriented services: Some implications for routine practice. *Criminal Behaviour and Mental Health*, *23*, 230-240.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, *56*, 63-75.

Myers, R. H. (1990). *Classical and modern regression with applications*. Belmont, CA: Duxbury Press.

Nester, P. (2002). Mental disorder and violence: personality dimensions and clinical features. *American Journal of Psychiatry*, *159*, 173-178. doi: 10.1176/appi.ajp.159.12.173

Nicholls, T., Brink, J., Desmarais, S., Webster, C., & Martin, M. (2006). The Short-Term Assessment of Risk and Treatability (START): A prospective validation study in a forensic psychiatric hospital. *Assessment*, *13*, 313-327. doi: 10.1177/1073191106290559

Ontario Review Board (2012). *Annual Report, Fiscal Period April 1, 2010 to March 31, 2011*. Retrieved from http://www.orb.on.ca/scripts/en/resources/Annual%20Report%202010-2011-EN.pdf

Ontario Review Board (2015). *Annual Report, Fiscal Period April 1, 2014 to March 31, 2015*. Retrieved from http://www.orb.on.ca/scripts/en/resources/Annual%20Report%202014-2015-EN.pdf

Penney, S., McMaster, R., & Wilkie, T. (2014). Multirater reliability of the Historical, Clinical, and Risk Management-20. *Assessment, 21*, 15-27.

Penney, S. R., Morgan, A., & Simpson, A. I. (2016). Assessing illness-and non–illness-based motivations for violence in persons with major mental illness. *Law and Human Behavior*, *40*, 42-49.

Personal Health Information Protection Act, 2004. S.O. 2004, c.3, Schedule A. Retrieved from http://www.e-laws.gov.on.ca/html/statutes/english/ elaws_statutes_04p03_e.htm. Accessed June 2016.

Quanbeck, C. (2006). Forensic psychiatric aspects of inpatient violence. *Psychiatric Clinics of North America*, *29*, 743-760.

Quinsey, V., Harris, G., Rice, M., & Cormier, C. (1998). *Violent offenders: Appraising and managing risk*. Washington: American Psychological Association.

Quinsey, V., Harris, G., Rice, M., Cormier, C. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.

Reimann, B., & Nussbaum, D. (2011). Predicting seclusion in a medium secure forensic inpatient setting. *International Journal of Forensic Mental Health*, *10*, 150-158.

Rice, M., & Harris, G. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and human behavior*, *29*, 615-620. doi: 10.1007/s10979-005-6832-7

Robinson, C., & Schumacker, R. (2009). Interaction effects: centering, variance inflation factor, and interpretation issues. *Multiple Linear Regression Viewpoints, 35*, 6-11.

Rogers, R. (2000). The uncritical acceptance of risk assessment in forensic practice. *Law and Human Behavior, 24*, 595-605.

Rutter, M. (1979). Protective factors in children's responses to stress and disadvantage. *Annals of the Academy of Medicine, Singapore, 8*, 324.

Rutter, M. (1985). Resilience in the face of adversity. Protective factors and resistance to psychiatric disorder. *British Journal of Psychiatry, 147*, 598-611.

Ryba, N. (2008). The other side of the equation: Considering risk state and protective factors in violence risk assessment. *Journal of Forensic Psychology Practice, 8*, 413-423- doi: 10.1080/15228930802199382

Sameroff, A., Bartko, W., Baldwin, A., Baldwin, C., & Seifer, R. (1998). *Family and social influences on the development of child competence.* In M. Lewis & C. Feiring (Eds.), Families, risk, and competence. Mahwah, NJ: Erlbaum.

Schuldberg, D., Quinlan, D., Morgenstern, H., & Glazer, W. (1990). Positive and negative symptoms in chronic psychiatric outpatients: Reliability, stability, and factor structure. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*, 262-268.

Seligman, M. (1998). The president's address. *American Psychologist, 54*, 559-562.

Seligman, M. (2002). Positive psychology, positive prevention, and positive therapy. In C.R. Snyder & S.J. Lopez (Eds.), *Handbook of positive psychology* (pp.3-9). New York: Oxford University Press.

Serin, R. (2007). The Dynamic Risk Assessment Scale for Offender Re-Entry (DRAOR). *Unpublished scale. Carleton University, Ottawa, Ontario*.

Serin, R., & Lloyd, C. (2009). Examining the process of offender change: The transition to crime desistance. *Psychology, Crime & Law, 15*, 347-364.

Siegert, R., Ward, T., Levack, W., & Mcpherson, K. (2007). A Good Lives Model of clinical and community rehabilitation. *Disability and Rehabilitation*, *29*, 1604-1615. doi:10.1080/09638280701618794.

Simons, D., McCullar, B., & Tyler, C. (2008). The utility of the self-regulation model to re-integration planning. In *27th Annual Association for the* Treatment *of Sexual Abusers Research and Treatment Conference, Atlanta, GA.*

Simpson, A., Penney, S., Seto, M., Crocker, A., Nicholls, T., & Darby, P. (2014). *Changing characteristics of the review board population in Ontario: A population-based study from 1987-2012*. Report for the Ministry of Health and Long Term Care, Toronto, Ontario.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*, 72-101

Stouthamer-Loeber, M., Loeber, R., Farrington, D. P., Zhang, Q., van Kammen, W., & Maguin, E. (1993). The double edge of protective and risk factors for delinquency: Interrelations and developmental patterns. *Development and Psychopathology*, *5*, 683–701.

Sturidsson, K., Haggard-Grann, U., Lotterberg, M., Dernevik, M., & Grann, M. (2004). Clinicians' perceptions of which factors increase or decrease the risk of violence among forensic out-patients. *International Journal of Forensic Mental Health*, *3*, 23-36.

Sutherland, A., Merrington, S., Jones, Baker, K., & Roberts, C. (2005). Youth Justice Board, & United Kingdom. *Role of Risk and Protective Factors.* Retrieved from https://www.youth-justice-board.gov.uk

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240,* 1285–1293.

Tabachnick, B., & Fidell, L. (2000). Using multivariate statistics, 4th edition. Allyn & Bacon.

Tamatea, A., & Wilson, N. (2009). *Dynamic risk assessment for offender re-entry (DRAOR): A pilot study.* New Zealand: Department of Corrections.

Thayer, J. (2002, April). *Stepwise regression as an exploratory data analysis procedure*. Paper session presented at the meeting of the Annual Meeting of the American Educational Research Association, New Orleans.

Trulson, C., DeLisi, M., & Marquart, J. (2009). Institutional misconduct, delinquent background, and rearrest frequency among serious and violent delinquent offenders. *Crime & Delinquency, 57*, 709-731.

Ttofi, M., Farrington, D., Piquero, A., Lösel, F., DeLisi, M., & Murray, J. (2016). Intelligence as a protective factor against offending: A meta-analytic review of prospective longitudinal studies. *Journal of Criminal Justice, 45*, 4-18.

Turner, D., Rettenberger, M., Yoon, D., Klein, V., Eher, R., & Briken, P. Risk assessment in child sexual abusers working with children. *Sexual Abuse: A Journal of Research and Treatment*, 1-25. doi: 10.1177/1079063214564390

Ullrich, S., & Coid, J. (2012). Protective factors for violence among released prisoners-effects over time and interactions with static risk. *Journal of Consulting and Clinical Psychology, 79*, 381-390.

Verdun-Jones, S. (1994). The insanity defence in Canada: Setting a new course. *International Journal of Law and Psychiatry, 17*, 175-189.

Vincent, G., Guy, L., Fusco, S., & Gershenson, B. (2011). Field reliability of the SAVRY with juvenile probation officers: Implications for training. *Law and Human Behavior, 36*, 225–236.

Vojt, G., Thomson, L., & Marshall, L. A. (2013). The predictive validity of the HCR-20 following clinical implementation: does it work in practice? *Journal of Forensic Psychiatry & Psychology*, *24*, 371-385.

Walters, G. (2011). *Crime in a psychological context: From career criminals to criminal careers*. Sage.

Ward, T. (2002a). Good lives and the rehabilitation of offenders. Promises and problems. *Aggression and Violent Behavior, 7*, 513-528.

Ward, T. (2002b). The management of risk and the design of good lives. *Australian Psychologist, 37*, 172-179.

Ward, T., & Brown, M. (2004). The Good Lives Model and conceptual issues in offender rehabilitation. *Psychology, Crime & Law, 10*, 243-257.

Ward, T., Mann, R. E., & Gannon, T. A. (2007). The Good Lives Model of offender rehabilitation: Clinical implications. *Aggression and Violent Behavior*, *12*, 87–107

Ward, T., & Stewart, C. (2003). Criminogenic needs and human needs: A theoretical model. *Psychology, Crime & Law, 9*, 125-143.

Ward, T., Yates, P. M., & Willis, G. M. (2012). The Good Lives Model and the Risk Need Responsivity model a critical response to Andrews, Bonta, and Wormith (2011). *Criminal Justice and Behavior*, *39*, 94-110.

Webster, C., Douglas, K., Eaves, D., & Hart, S. (1997). *HCR-20. Assessing the risk of violence*. Version 2. Burnaby, British Columbia: Simon Fraser University and Forensic Psychiatric Services Commission of British Columbia.

Webster, C., Eaves, D., Douglas, K., & Wintrup, A. (1995). *The HCR-20 scheme: The assessment of dangerousness and risk.* Vancouver: Simon Fraser University and British Columbia Forensic Psychiatric Services Commission.

Webster, C., Martin, M., Brink, J., Nicholls, T., Middleton, C. (2004). *Short-Term Assessment of Risk and Treatability (START)* (Version, 1.1). Coquitlam, BC, Canada: British Columbia Mental Health and Addiction Services.

Wechsler D. (1999). *Weschler Abbreviated Scale of Intelligence* (WASI). London: Harcourt Assessment.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence–Second Edition Manual*. Bloomington, MN: Pearson.

Wilson, C., Desmarais, S., Nicholls, T., & Brink, J. (2010). 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, 9,* 282-293. doi: 10.1080/14999013.2010.534694

Winko v. British Columbia. (1999). *Forensic Psychiatric Institute,* S.C.C. File 25856.

Wong, P. T. (2011). Positive psychology 2.0: Towards a balanced interactive model of the good life. *Canadian Psychology/Psychologie Canadienne, 52*, 69-81.

Worling, J. R. (2013). *Desistence for Adolescents who Sexually Harm (DASH-13).* Unpublished manuscript. University of Toronto, Canada.

Worling, J., & Curwen, T. (2001). Estimate of risk of adolescent sexual offense recidivism (ERASOR; Version 2.0). *Juveniles and children who sexually abuse: Frameworks for assessment*, 372-397.

Yates, P. M. (2013). Treatment of sexual offenders: Research, best practices, and emerging

models. *International Journal of Behavioral Consultation and Therapy*, *8*, 1-7.

Yoon, D., Spehr, A., & Briken, P. (2011). Structured assessment of protective factors: A German pilot study in sex offenders. *The Journal of Forensic Psychiatry & Psychology*, *22*, 834-822.

Yoshikawa, H. (1994). Prevention as cumulative protection: Effects of early family support and education on chronic delinquency and its risks. *Psychological Bulletin*, *115*, 128-154.

Zawadzki, J. (2014). SAPS/SANS interview guide (unpublished). Toronto, Ontario.

Zeng, G., Chu, C., & Lee, Y. (2015). Assessing protective factors of youth who sexually offended in Singapore preliminary evidence on the utility of the DASH-13 and the SAPROF. *Sexual Abuse: A Journal of Research and Treatment*, *27*, 91-108.

Cases

R. v Chaulk (1990). 3 S.C.R. 1303

R. v M'Naghten. (1843). 10 Cl. and F. 200, 8 Eng. Rep.718.

R. v Swain. (1991). 63 C.C.C. (3d) 481 (S.C.C.)