ASSESSING RISK FOR INPATIENT VIOLENCE ON HIGH-SECURITY FORENSIC PSYCHIATRIC UNITS

A Dissertation Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Psychology

University of Saskatchewan

Saskatoon

By

Neil Robert Hogan

PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my dissertation.

DISCLAIMER

Reference in this dissertation to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes.

Requests for permission to copy or to make other uses of materials in this dissertation in whole or part should be addressed to:

Head of the Department of Psychology
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A5
Canada

OR

Dean
College of Graduate Studies and Research
University of Saskatchewan
107 Administration Place
Saskatoon, Saskatchewan S7N 5A2
Canada

ABSTRACT

While forensic psychiatric inpatient violence is a serious problem, research on risk assessment for this outcome is limited; the current research investigated the predictive validity of a number of structured risk/forensic instruments for inpatient violence. Research objectives included: 1) observing the profile of dynamic changes in violence risk detected by existing violence risk assessment instruments; 2) assessing whether existing violence risk assessment instruments could be used to assess risk for inpatient violence; 3) evaluating the contribution of dynamic risk measures to the prediction of inpatient violence; and 4) assessing the relationship between dynamic changes in risk and inpatient violence. Instruments included: the Historical Clinical Risk Management 20 - Version 3 (HCR-20^{V3}), the Psychopathy Checklist Revised (PCL-R), the Short-Term Assessment of Risk and Treatability (START), the Revised Violence Risk Appraisal Guide (VRAG-R), and the Violence Risk Scale (VRS). Two studies were conducted on a maximum-security forensic psychiatric unit at Alberta Hospital Edmonton. Study 1 was a pseudo-prospective archival investigation (n = 99), while Study 2 was a prospective investigation (n = 19); all risk assessment scores were based on information available in institutional files.

Instruments designed to capture dynamic/clinical risk variables (HCR-20^{V3}, START, VRS) detected dynamic changes in risk in this setting over longer follow-ups (i.e., between admission and discharge), but not over shorter follow-ups (i.e., 28 day periods). Predictive validity analyses indicated that specialized measures designed to capture relevant dynamic/clinical variables were significant predictors of inpatient violence; instruments that were not designed for this purpose (PCL-R and VRAG-R), did not demonstrate predictive validity for inpatient violence. Dynamic measures consistently demonstrated incremental predictive validity for inpatient violence, beyond the static measures. Additionally, change scores demonstrated incremental relationships with decreased inpatient violence, beyond pretreatment scores. Put another way, positive risk change was associated with decreased violence over the course of the patients' stays in hospital.

Reliable and valid risk assessments are a necessary component of effective offender programming (Risk-Need-Responsivity Model) and the current results indicated that valid violence risk assessments for forensic psychiatric inpatient violence are possible. Implications for clinical practice and the reduction/mitigation of inpatient violence are discussed.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to those who made this dissertation project possible. Dr. Mark Olver, my doctoral research supervisor, has proven to be an exceptional mentor. While continuing to distinguish himself as a leading researcher through his own merit, his focus on his students' success has never wavered. I am truly grateful for his guidance and support throughout my doctoral program, and for the opportunity to have a professional role model of such character. I would like to thank my dissertation committee, Dr. Nicholaichuk, Dr. Woods, and Dr. Wormith, for sharing their insight and experience, which elevated the quality of this document. I would also like to thank my external examiner, Dr. Nicholls, for offering her own valuable insight into the project.

I owe special thanks to Dr. Andrew Haag, who generously offered his support to this research from the beginning, and who remained generous with his knowledge and time (and office space) until the end. The other staff at AHE, from management to clinical staff and administrative support, are also deserving of my thanks for facilitating this research in a variety of ways. I would particularly like to thank Gabriela Corabian and Erika van Husen, who offered invaluable assistance through coding, file collection, and a variety of other tasks.

To my parents: I cannot possibly thank you enough for the numerous ways, large and small, that you have supported me in reaching this goal; thank you for the invaluable gift of perspective. To Kieran, Jessica, Victor, Laura, Breanne, and Matt: thank you for your encouragement and for helping me recharge during my trips back home. To Gabi: thank you for your ever-present support, encouragement, and belief. I would also like to thank the close friends I made in Saskatoon for greatly improving my graduate school experiences, and the old friends at home, for numerous helpful reminders of a world outside of graduate school.

Finally, I would like to gratefully acknowledge the financial support received from the Social Sciences and Humanities Research Council, in the form of Joseph-Armand Bombardier Canada Graduate Scholarships Masters (766-2011-0092) and Doctoral (767-2012-1659) Awards. Additionally, I would like to thank the University of Saskatchewan Centre for Forensic Behavioural Sciences and Justice Studies for supporting the research with a student award.

DEDICATION

This dissertation is dedicated to my family and friends, including those that I have lost.

Your absence has gone through me

Like thread through a needle.

Everything I do is stitched with its color.

W. S. Merwin

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

Chapter 1: ASSESSING RISK FOR INPATIENT VIOLENCE ON HIGH-SECURITY FORENSIC PSYCHIATRIC UNITS

1.1 General Introduction

Inpatient violence in psychiatric settings is a widespread problem, although it is particularly prevalent among forensic psychiatric populations (Hill, Rogers, & Bickford, 1996; McDermott, Edens, Quanbeck, & Busse, 2008). Patients and professionals exposed to inpatient violence may suffer a variety of physical (Bowers, Allan, Simpson, Nijman, & Warren, 2007; Daffern & Howells, 2002; Johnson, 2004) and psychological (Inoue, Tsukano, Muraoka, Kaneko, & Okamura, 2006; Johnson, 2004; Rippon, 2000) consequences. Despite these serious consequences, inpatient violence in forensic settings has received relatively little research attention, particularly when compared to violence in the community. Fortunately, previous research addressing other forms of criminal and violent behaviour can be used to guide efforts to address this problem. Specifically, treatments based on the Risk-Need-Responsivity model (RNR; Andrews, Bonta, & Hoge, 1990) have received strong research support (e.g., Andrews & Bonta, 2010; Hanson, Bourgon, Helmus, & Hodgson, 2009). Treatments based on this model are characterized by the following: 1) according to the Risk principle, the highest risk offenders receive the most services; 2) according to the Need principle, services should be targeted to address criminogenic needs (e.g., social, psychological, and emotional factors associated with risk for criminal behaviour); and 3) according to the Responsivity principle, effective interventions are tailored to the individual's unique characteristics (Andrews, Bonta, & Wormith, 2011). With these principles in mind, the current research was designed to add to the limited research on forensic psychiatric inpatient violence, by investigating violence risk assessment among forensic psychiatric inpatients.

1.2 Violence Risk Assessment

1.2.1 The Development of Risk Assessment Practices

The formal assessment of risk for violent recidivism involves judgements about the likelihood of uncertain outcomes (Hanson, 2009). Over time, the nature and bases of these judgements have evolved from relying exclusively on unstructured professional opinions towards increasingly systematic and empirically informed approaches. This evolution has been conceptualized (Bonta, 1996; Bonta & Andrews, 2007) as progressing through successive generations of risk assessment "technology" (Quinsey, 2009, p. 15). The following discussion

will focus on the evolution of risk assessment practices in general, and thus it applies to techniques and instruments designed to assess risk for general recidivism (i.e., both violent and non-violent recidivism) as well as those designed specifically to assess risk for violent recidivism.

1.2.1.1 First Generation. Prior to the 1970s, risk assessments were generally comprised of the unstructured clinical judgements of professionals (Bonta & Andrews, 2007; Campbell, French, & Gendreau, 2009). The problems associated with employing unstructured approaches in the prediction of human behaviour, in general terms, were discussed by Meehl (1954) in his now classic book, *Clinical Versus Statistical Prediction*. Meehl reviewed existing empirical literature comparing two methods of combining assessment data to predict human outcomes: clinical judgement and statistical/mechanistic techniques. From a total of 20 studies dealing with such diverse subject matter as psychiatric prognosis, vocational achievement, and criminal recidivism, Meehl identified at least 19 as having indicated that statistical techniques were equal to or better than clinical judgements at predicting behavioural outcomes. Nearly half a century later, Grove, Zald, Lebow, Snitz, and Nelson (2000) conducted a meta-analysis focused on the same issue and found similar results to those cited by Meehl. Of 136 studies published between 1936 and 1989, 128 indicated that mechanical/statistical techniques were equal to or better than clinical judgements at predicting health-related and behavioural outcomes.

Other studies have challenged the effectiveness of clinical judgement for the assessment of violence risk in particular. For example, a number of studies following the landmark Baxstrom case in New York (Steadman & Cocozza, 1974) seriously undermined the credibility of contemporary violence risk assessment practices. The Baxstrom case resulted in the transfer of patients who had previously been identified as high risk for violence from high-security forensic hospitals to civil psychiatric facilities. Despite their previous status, Steadman and Keveles (1972) reported that of 121 Baxstrom patients released to the community, only one was convicted of a violent offense (average follow-up time of approximately 2.5 years). Other research also challenged the validity and nature of *expert* clinical judgements of violence risk. For example, Quinsey and Ambtman (1979) found that expert psychiatrists' and untrained school teachers' assessments of the likelihood of assaultive behaviour among mentally disordered offenders were similar (rs = .55 to .83, all correlations significant at $p < 0.01$), regardless of the type of data provided to inform these judgements. The results also demonstrated low levels of

agreement among psychiatrists' decisions to release patients. These and similar findings led to the development of the next generation of violence risk assessment technology.

1.2.1.2 Second Generation. In contrast with unstructured clinical judgement, second generation risk assessment involves standardization, based on statistical relationships among measurable variables. Notably, Burgess (1928) had demonstrated the utility of this concept years before Meehl's criticism of clinical judgement, by studying relationships between risk variables and recidivism among parolees. However, despite the fact that Burgess' method has now been recognized (Bonta, Harman, Hann, & Cormier, 1996) as having influenced subsequent researchers, second generation risk assessment did not truly come to the fore until years later. As an example of second generation research, Nuffield (1982) analyzed data obtained from a sample of Canadian offenders released to the community and found a relationship between rearrests and scores on the Statistical Information on Recidivism (SIR) scale. These scores were based on a combination of risk variables, such as young age at first conviction and number of previous convictions for a violent offence. Bonta and colleagues (1996) later validated the SIR with a sample of Canadian offenders and found a small but significant relationship between SIR scores and violent recidivism. Another example of second generation risk assessment is the Violence Risk Appraisal Guide (VRAG; Harris, Rice & Quinsey, 1993), which was statistically developed for the prediction of violent recidivism among mentally disordered offenders and which was found to discriminate between violent recidivists and non-recidivists. Harris, Rice, and Cormier (2002) later replicated the early results, obtaining a large effect size for the relationship between VRAG scores and violent recidivism and demonstrating the superiority of VRAG scores over clinical judgement.

While researchers have empirically demonstrated relationships between violent recidivism and scores produced by second generation risk assessment instruments, some of the limitations of these instruments may be identified through the lens of a clinician or professional charged with risk reduction. Actuarial tools rely heavily on historical and therefore largely static (i.e., unchangeable) risk factors. As a result, they are considered to be poorly suited to the assessment of reductions in risk (Bonta & Andrews, 2007) and the identification of targets for treatment. Various authors (e.g., Andrews et al., 1990; Douglas & Skeem, 2005; Olver, Wong, Nicholaichuk, & Gordon, 2007; Yang, Wong, & Coid, 2010), have identified a need to move beyond risk assessment for its own sake, towards effective interventions and reductions in rates

of recidivism. As Hart (1998) has argued, a completed risk assessment should not be the clinician's end goal.

1.2.1.3 Third Generation. Third generation risk assessment addressed some of these concerns through the selective addition of content to the empirical strengths of second generation instruments. In particular, the inclusion of dynamic or changeable variables in risk assessment allowed third generation instruments to identify targets for treatment and to detect reductions in risk. Third generation risk assessment instruments aim to capture dynamic risk factors, or *criminogenic needs* (Andrews et al., 1990), that when altered, are associated with corresponding changes in the probability of recidivism. In addition to incorporating criminogenic needs, third generation risk assessment instruments may also be differentiated from earlier generation instruments in that they are more likely to have a theoretical basis (Andrews, Bonta, & Wormith, 2006). Examples of third generation risk assessment instruments include the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995) and the Historical-Clinical-Risk Management Violence Risk Assessment Scheme-Version 2 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997). The LSI-R is based on the Risk-Need-Responsivity model (RNR; Andrews et al., 1990). It combines static risk factors like criminal history, along with criminogenic needs like procriminal attitudes, and has been shown to predict both general and violent recidivism among offenders (Gendreau, Goggin, & Smith, 2002). The HCR-20 is broken down into three domains: historical risk factors, such as previous violence; clinical or current risk factors, such as mental health symptoms; and risk management or potential future risk factors, such as problems with social support. This instrument has also been validated across a number of offender populations, with a meta-analysis by Yang and colleagues (2010) finding a moderate level of predictive accuracy.

1.2.1.4 Fourth Generation. This most recent generation of risk assessment includes instruments that were designed for use over the entire duration of an offender's involvement with the criminal justice system (Campbell et al., 2009). According to Andrews and colleagues (2006), fourth generation instruments provide structured means of monitoring cases and facilitating adherence to principles of effective offender rehabilitation. Two examples of fourth generation instruments are the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004) and the Violence Risk Scale (VRS; Wong & Gordon, 1999-2003); like the LSI-R, both the LS/CMI and the VRS are based on RNR principles (Andrews et

al., 1990). Wormith, Olver, Stevenson, and Girard (2007) assessed the risk assessment function of the LS/CMI and found that it was able to predict future violence. Similarly, Wong and Gordon (2006) have found evidence supporting the predictive validity of both the static and the dynamic items on the VRS. Thus, it appears that while building in these additional functions, these newer instruments have maintained the gains in predictive validity developed through the previous generations of risk assessment practices.

1.2.1.5 Further Technical Considerations in Risk Assessment Classification. In addition to this multi-generational conceptualization which is based primarily on content, risk assessment instruments may also be classified using technical criteria. Contemporary instruments differ in the manner in which they are used to produce a final determination of risk. Some instruments, like the VRAG and the LSI-R, combine risk information using algorithms. After the assessor has scored each of the items that comprise the instrument, this information is combined in a predetermined and structured way to arrive at a final determination of risk. An alternative to this approach is the Structured Professional Judgement (SPJ; Douglas, Blanchard, Guy, Reeves, & Weir, 2010) technique associated with tools like the HCR-20. Under the SPJ approach, the assessor considers each of the empirically identified risk factors that constitute a given instrument individually, and then makes a final conclusion regarding overall risk based on the balance of information. Considerable variability exists among risk assessment techniques and practices; the instruments most often employed by violence risk assessors (Archer, Buffington-Vollum, Stredny, & Handel, 2006) come from different generations and employ different approaches.

1.2.2 Quantifying the Accuracy of Contemporary Violence Risk Assessment

The preceding discussion of evolving risk assessment practices briefly cited evidence for *relationships* between instruments and recidivism, as well as conclusions regarding various instruments' abilities to *predict* violence. Additional information is required if one is to draw informed conclusions regarding what contemporary risk assessment instruments can and cannot do. Any statement regarding the utility of a risk assessment tool, or a particular individual's risk for violence for that matter, requires important qualifications. Examples of relevant questions that must be addressed include: what is the outcome being predicted (e.g., formal charges, convictions, specific behaviours, etc.), how is risk quantified or presented (e.g., a relative low-, moderate-, or high-likelihood appraisal versus recidivism rates in a normative sample), to what

populations do the results apply and in what contexts, over what time period does the appraisal apply (e.g., days, months, or years), and what are the relevant base rates of the outcome (i.e., how often does it occur in the population)?

1.2.2.1 Difficulties in the Assessment of Risk for Violence. Researching violence risk assessment practices has proven to be a difficult task for a number of reasons. For example, violent recidivism does not lend itself easily to experimental study. As Hanson (2009) pointed out, in practice, risk assessment decisions (i.e., an independent variable) tend to influence violent outcomes (i.e., a dependent variable). If offenders are deemed to be at a high risk for violent recidivism, it follows that they may be allowed fewer opportunities to recidivate (Quinsey, Harris, Rice, & Cormier, 2006) due to the imposition of increased sanctions or supervision. Even among offenders with the opportunity to recidivate, identifying and defining appropriate, valid, and reliable outcome variables remains a difficulty (Yang et al., 2010). Definitions of violence vary and operationalized outcome variables in the literature range from formally recorded convictions to self-report data. As a result, research findings can be difficult to combine and compare.

Historically, the base rate problem proved to be a particularly obstinate impediment to meaningful risk assessment. Base rates (Meehl & Rosen, 1955) refer to the probabilities of events in given populations and may be established through such means as historical data or actuarial tables. For example, given a base rate of 50% for "heads" coming up in coin tosses, one could predict a result of "heads" for all future tosses and reasonably expect to be correct approximately 50% of the time. To be useful, a risk assessment instrument should improve on this form of prediction. This matter is complicated further by the fact that violence risk assessment instruments, like most psychometric instruments and diagnostic practices, are imperfect; as Hanson (2009) aptly put it: "causal mechanisms for human behavior are sufficiently complex that it is impossible to reason through all the possibilities" (p. 172). Thus, a tool used to identify future recidivists will produce four outcomes to varying degrees (Quinsey et al., 2006): 1) correct identifications of recidivists, also called hits or true-positives; 2) correct identifications of non-recidivists, or true-negatives; 3) failures to identify future recidivists, also called misses or false-negatives; and 4) the incorrect identification of non-recidivists as recidivists, also called false-positives or false alarms. In the case of a phenomenon with low base rates, as was the case with the rate of violent recidivism among the Baxstrom patients

(Steadman & Keveles, 1972), even a low *rate* of false-positives results in an inordinate number of cases being incorrectly identified as at-risk (Megargee, 1976).

Over time, research conducted since the Baxstrom case has shown that the base rates of violence among various populations (e.g., previously violent offenders) are high enough to allow for meaningful risk assessment (Quinsey, 2009). Additionally, researchers have applied methods of statistical analysis that are less affected by base rates, such as the Receiver Operating Characteristic (ROC; Andrews & Bonta, 2010; Humphreys & Swets, 1991; Quinsey et al., 2006). Nonetheless, base rates remain an important consideration in the development and evaluation of risk assessment instruments and practices.

1.2.2.2 Empirical Evidence for Accuracy of Violence Risk Assessment. According to Quinsey (2009), current actuarially-based risk assessment practices, "are the most accurate known" (p. 17), but this is a relative statement and further elaboration is warranted. As mentioned previously, a discussion of the accuracy of a risk assessment procedure should explicitly identify such matters as the applicable populations, settings, time-scales, and outcomes. With these caveats in mind, there is a considerable body of research across diverse settings which can provide quantitative estimates of the accuracy of violence risk assessment practices.

To begin, Hanson (2005) argued that to understand the empirical research in a given area, one should first review existing meta-analyses. Meta-analyses provide quantitative summaries of research evidence and indications of the "magnitude and direction of empirical relationships" (p. 215). Because they combine the results of multiple studies, meta-analyses' results are less likely to be influenced by anomalous or atypical circumstances (i.e., related to research design, sample characteristics, unusual contextual factors, etc.) than the results of individual studies. Thus, the relationships identified by meta-analyses may also be considered more likely to generalize to other populations than those found in individual studies. Results of previous meta-analyses will now be discussed, to provide estimates of the accuracy of violence risk assessment practices.

Bonta, Law, and Hanson (1998) analyzed the predictors of violent recidivism in studies of offenders with and without mental disorders. They disaggregated existing tools and grouped predictors into four categories: personal demographic variables (e.g., young age), criminal history variables (e.g., history of violent behaviour), deviant lifestyle variables (e.g., poor work adjustment), and clinical variables (e.g., antisocial personality disorder diagnosis). Based on the

mean weighted *r* values, three of the four groupings of predictors were found to be positively related to violent recidivism. They are listed here in order from largest mean effect size to smallest: criminal history (mean weighted $r = .15$), personal demographics (mean weighted $r = .12$), and deviant lifestyle (mean weighted $r = .08$). When analyzed as a group, the relationship between the clinical variables and violent recidivism was not significant (mean weighted $r = -.03$). However, the mean effect size for antisocial personality alone indicated that this diagnosis was significantly related to violent recidivism (mean weighted $r = .18$). Finally, the relationship between violent recidivism and the empirically validated objective risk assessment scales, considered as intact wholes, was also significant (mean weighted $r = .30$).

A decade on from Bonta and colleagues' study, Campbell and colleagues (2009) conducted a meta-analysis to assess the relative abilities of a number of risk instruments and psychometric measures to predict violence. In addition to a number of commonly studied second- and third-generation risk instruments (HCR-20, LSI/LSI-R, SIR, and VRAG), these researchers also included three variants of Hare's psychopathy measures: the Psychopathy Checklist (PCL; Hare, 1980), the Psychopathy Checklist-Revised Second Edition (PCL-R; Hare, 2003), and the Psychopathy Checklist: Screening Version (PCL:SV; Hart, Cox, & Hare, 1995). For the prediction of violent recidivism, there was little difference observed among the performance of the various measures, as demonstrated by mean effect sizes (M_r) ranging from .24 to .27 and overlapping 95% confidence intervals. However, Campbell and colleagues noted that for violent recidivism, their data provided some support for the use of dynamic over static risk predictors, third-generation over second-generation instruments, and file-review plus interview methods over other forms of risk assessment. Notably, their analysis of institutional violence produced somewhat different results, and these will be discussed in a subsequent section of this paper.

Yang, Wong, and Coid (2010) conducted another meta-analysis, with a focus on more recent studies of structured forensic instruments. They included a number of the same tools as Campbell and colleagues, such as the PCL-R, PCL:SV, VRAG, HCR-20, SIR, and LSI-R. Additionally, Yang and colleagues included the Offender Group Reconviction Scale (OGRS; Copas & Marshall, 1998), the Risk Matrix 2000 for Violence (RM2000V; Thornton, 2007), and the VRS. Effect size estimates for these tools using Cohen's *d* ranged from .55 to .79, and few substantive differences among instruments were noted, which led the authors to conclude that for

the purposes of violence risk prediction, the instruments were "essentially interchangeable" (p. 759).

Fazel, Singh, Doll, and Grann (2012) conducted a systematic review and meta-analysis of the results of a number of studies of structured risk assessment for violent and antisocial behaviour. These authors reported a mean AUC of .72 for violence risk assessment instruments, which included the HCR-20, VRAG, Spousal Assault Risk Assessment Guide (SARA; Kropp, Hart, Webster, & Eaves, 1995), and the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2002). The authors cautioned that it would be inappropriate to use any risk assessment instrument alone to make important legal and risk management decisions, but their results nonetheless indicated that these instruments may contribute valuable information to such decisions.

For the purposes of the current review, these empirical findings lead to two important conclusions. First of all, the magnitudes of the effect sizes are respectable (Rice & Harris, 2005) and indicate that structured risk assessment instruments in general are accurate enough to be useful in the assessment of risk for violence (Hanson, 2009; Quinsey, 2009). Second of all, they suggest that there is no gold-standard for assessing risk for violence; many risk instruments are correlated with one another and are able to achieve similar rates of success (e.g., Yang et al., 2010). This conclusion is consistent with a study of general criminal recidivism conducted by Kroner, Mills, and Reddon (2005), who found that randomly sorted combinations of the items of four common risk assessment instruments predicted antisocial behaviour as well as the original instruments. Based on these results, it is difficult to argue against the conclusion drawn by authors such as Douglas (2014) and Yang and colleagues (2010), that choices among risk instruments should be based on the intended uses of the tools (e.g., pure risk assessment, treatment planning, or the measurement of treatment progress), the population being assessed, and the setting. Furthermore, continued efforts to improve upon existing methods of violence risk assessment are warranted.

1.2.2.3 Some Recent Developments in Violence Risk Assessment. As the field continues to develop, new instruments are emerging and existing instruments are being revised. For example, since the three major meta-analyses of structured schemes described above were published (Campbell et al., 2009; Fazel et al., 2012; Yang et al., 2010), both the VRAG and the HCR-20 have been revised, and these two instruments were found to be among the most

commonly used risk assessment instruments in a survey of forensic psychologists (Archer et al., 2006). The following is a review of the early research on updated versions of these two major risk assessments schemes: the Historical-Clinical-Risk Management-20 Version 3 (HCR-20^{V3}) and the Violence Risk Appraisal Guide - Revised (VRAG-R).

1.2.2.3.1 ***HCR-20 Version 3***. In keeping with the previous iterations, the HCR-20^{V3} (Douglas, Hart, Webster, & Belfrage, 2013) is based on the SPJ approach to violence risk assessment. Items reflect: historical risk factors, such as violence history; clinical factors, such as problems associated with insight into violence risk or mental disorder; and risk management factors, such as problems with future plans. Some new features of the most recent version will be described briefly here, as they are particularly relevant for the purposes of comparing its predictive validity to that of its predecessors (Version 1, Webster, Eaves, Douglas, & Wintrup, 1995; Version 2, Webster et al., 1997). First, the authors have added two new SPJ summary statements to the overall assessment of risk, pertaining to imminent risk for violence and to potential for serious harm, respectively. The addition of the Imminent Violence summary statement appears to be particularly relevant to the subject of the current research program, given the intended emphasis on differentiating between relatively short-term risk for inpatient violence, and longer term risk for community recidivism. Second, the authors have explicitly separated the rating of individual risk factors into two dimensions, by encouraging assessors to consider both Presence and Relevance scores. This latter change was made with the intention of facilitating assessment practices that go beyond the formulaic aggregation of items, towards a more meaningful formulation of risk for the individual (Douglas, 2014); when scoring the relevance of items, considerations might include, for example, a behavioural analysis of previous violent behaviour. Thus, while previous research evaluating the overall predictive validity of the HCR-20 scheme generally considered the overall SPJ summary risk rating or a single total score, research examining the most recent version may consider up to three SPJ summary ratings, and two total scores (Presence and Relevance).

The few published studies focusing on the HCR-20^{V3} have provided early indications of the reliability and validity of the instrument. Kötter and colleagues (2014) examined the interrater reliability of a German translation of the HCR-20^{V3} and found that after a two-day training workshop, coders demonstrated good consistency for most individual items, with average intraclass correlation coefficients (ICCs) for items on the individual scales ranging from

.65 to .73. Douglas and Belfrage (2014) evaluated interrater reliability with a Swedish translation, and produced ICC values for scale totals ranging from .69 to .94. Additional interrater reliability analyses conducted by de Vogel, van den Broek, and de Vries Robbe (2014) and Smith, Kelley, Rulesh, Sörman, and Edens (2014) produced similar results, generally supporting the interrater reliability of the HCR-20^{V3} overall. Given the extensive research base supporting the previous version of the HCR-20 scheme (Douglas et al., 2014), researchers have also examined the continuity between the HCR-20^{V3} and the HCR-20 Version 2. Published studies examining the concurrent validity of the two tools have indicated that they are high correlated, with reported *r* values of: .85 to .90 (Douglas & Belfrage, 2014), .93 (de Vogel et al., 2014), and .58 (Bjørkly, Eidhammer, & Selmer, 2014). Finally, a limited number of studies have also reported findings of predictive validity analyses. A pilot study conducted by de Vogel and colleagues (2014) demonstrated strong predictive validity with regard to violence, with AUCs of .77, .75, and .67, corresponding to follow-up periods of one, two, and three years, respectively. Doyle and colleagues (2014) reported AUCs for the prediction of violence of .73 and .70, corresponding to follow-up periods of six months and one year, respectively. Strub, Douglas, and Nicholls (2014) evaluated the predictive validity of both Presence and Relevance total scores for violence, among combined samples of civil psychiatric and offender populations. Across follow-up periods of four to six weeks, and six to eight months, the various scores produced AUCs ranging from .68 to .78. Overall, while these results may be considered preliminary, the existing HCR-20^{V3} research has demonstrated support for the instrument's psychometric properties, including predictive validity for violence.

1.2.2.3.2 *VRAG-R*. Rice, Harris, and Lang (2014) published an article describing the process of revising the VRAG, which involved attempts to improve on the predictive accuracy and useability of the actuarially derived instrument. Substantive changes included: the removal of items pertaining to the offenders' psychopathology (i.e., diagnoses of schizophrenia and personality disorders) and others related to their victims (i.e., gender of, and injury to victim); the substitution of the PCL-R Facet 4 score for the total score; the expansion of the item pertaining to substance abuse; and the addition of four new items, pertaining to violent and sexual offending history, conduct disorder, and periods of incarceration. At the time of writing, the current author was not aware of any additional published studies examining the psychometric properties of the newly developed VRAG-R, besides Rice and the other developers' (2014) own report of the

revision process. Nonetheless, the developers' findings demonstrated strong interrater reliability for the VRAG-R, with an ICC of .99 computed from the scores of two "experienced raters" (p. 8). Rice and colleagues VRAG-R results further demonstrated strong predictive accuracy for the prediction of violent recidivism, with an overall AUC of .76. While further study and replications are certainly warranted, these results suggest that the VRAG-R holds promise for the assessment of risk for violence.

1.2.3 Empirical Evidence of Dynamic Risk

As described above, existing third and fourth generation instruments were designed to capture dynamic risk factors and many of these instruments have demonstrated predictive validity for violent behaviour. However, while evaluating whether scores on risk assessment instruments can predict violence is informative in its own right, the extent to which ostensibly dynamic risk factors are genuinely dynamic is another empirical question altogether. To demonstrate that they can function as intended and provide incremental value beyond static predictors, dynamic risk factors require stricter tests. The identification of dynamic risk factors requires, at least, measurement of potential predictors at two time points, prior to the measurement of the outcome of interest (Andrews & Bonta, 2010; Andrews et al., 2006; Kraemer et al., 1997). It follows then, that to validate structured instruments for the assessment of dynamic violence risk, it is necessary to measure changes in risk scores, and to relate these changes to subsequent violent behaviour; as Douglas and Skeem (2005) argued, this general paradigm represents the essential test of a dynamic risk factor. Unfortunately, to date, this type of research has been limited in the field of violence risk assessment (Douglas & Skeem, 2005; Serin, Lloyd, Helmus, Derkzen, & Luong, 2013).

Some strong empirical evidence for the existence of genuinely dynamic violence risk factors has come from the study of sexually violent offenders. For instance, a study by Olver and colleagues (2007) evaluated the predictive validity of the dynamic items of the Violence Risk Scale – Sexual Offender Version (VRS-SO; Wong, Olver, Nicholaichuk, & Gordon, 2003) for sexual recidivism. The VRS-SO was adapted from the VRS, and the dynamic items on both instruments reflect a modified version of the transtheoretical model of change (TTM; Prochaska, DiClemente, & Norcross, 1992); if offenders are identified as having moved through the stages of change, reflecting risk-relevant cognitive and behavioural changes, corresponding changes are made to their VRS or VRS-SO risk scores. In Olver and colleagues' study, VRS-SO dynamic

change scores were computed by comparing the scores of 321 male offenders before and after their participation in a high-intensity sex offender treatment program. The results of Cox regression survival analyses indicated that dynamic change scores incrementally predicted sexual recidivism (e^B = 0.90) beyond measures of overall risk that were scored prior to the completion of treatment. More specifically, this result indicated that a one point increase in dynamic change scores (i.e., implied reduction in risk) was associated with a predicted 10% decrease in the hazard of sexual recidivism after controlling for baseline risk. Another study of the VRS-SO, conducted by Olver, Beggs Christofferson, Grace, and Wong (2014), included similar analyses with a combined sample of 539 sexually violent offenders in Canada and New Zealand. Controlling for static risk and pretreatment scores, dynamic change scores produced by the VRS-SO incrementally predicted both violent (e^B = .88) and sexual recidivism (e^B = .87). Put another way, a one point increase in dynamic change scores was associated with a predicted reduction of 12% in the hazard of violent recidivism and a 13% reduction in the hazard of sexual recidivism.

With regard to violent behaviour in the broader sense (i.e., including sexual and non-sexual violence), few rigorous evaluations of the dynamic nature of risk factors have been conducted to date. Serin and colleagues (2013) conducted a review of the existing literature on the relationship between dynamic risk and recidivism among forensic psychiatric inpatient populations, and found that very few studies had actually evaluated this relationship directly. They reported that based on their searches conducted in August/September of 2010, they found two studies that appeared to "hint" (p. 44) at a relationship between positive treatment-related changes and decreased recidivism among violent offenders, but also called for further research to be conducted on this important topic. Subsequent to Serin and colleagues' review paper, a limited number of additional studies have evaluated dynamic risk factors for violence. For example, Lewis, Olver, and Wong (2013) assessed the relationship between treatment related changes measured by the VRS and subsequent violent recidivism, among a sample of 150 male offenders. Dynamic risk changes were computed by comparing VRS scores obtained before and after offenders' participation in a treatment program designed to reduce aggressive behaviour. Results of Cox regression survival analyses indicated that dynamic change scores were incrementally predictive of violent recidivism, whether controlling for static risk (e^B = .91) or pretreatment total scores (e^B = .92). More recently, De Vries Robbé, de Vogel, Douglas, and Nijman (2015) published results of a study that assessed the predictive validity of dynamic

changes in risk, produced by the HCR-20 Version 2, for community recidivism. The sample included 108 male patients of two Dutch forensic psychiatric institutions. Results of binary logistic regression analyses indicated that the model including HCR-20 posttreatment scores demonstrated improved prediction of violent recidivism, beyond the model using only pretreatment scores, over a long-term follow-up (but not over a one-year follow-up period). Using Cox regression survival analyses, the investigators found that HCR-20 posttreatment scores improved the prediction model for violent recidivism beyond the use of pretreatment scores alone, while controlling for time at risk.

Overall, the results of existing studies seem to provide preliminary support for the contention that third and fourth generation risk instruments are capable of capturing genuinely dynamic risk factors, and using these risk factors to assess risk for violence. Even so, it is clear that further research is necessary to establish the predictive validity of individual instruments' respective dynamic components, for particular outcomes and settings of interest. This future research can benefit from the direction provided by the previous work.

1.2.4 Context Matters

Context has long been recognized as an important and complex influence on violence risk (Megargee, 1976). Different contexts vary with regard to the presence of various stressors, risk factors, and protective factors, including: levels of supervision and external controls, stability of housing arrangements, relative exposure to positive and negative peer influences, and access to formal treatment services. In keeping with this observation, violence risk assessment practices have been evaluated in different contexts.

1.2.4.1 Institutional Violence in Corrections. Institutional violence in corrections (i.e., violence within prisons and correctional facilities) provides a good example of this concept. In a paper analyzing failed attempts to decrease stabbing deaths and physical assaults against staff inside California prisons, Bidna (1975) speculated that institutional violence may be influenced by factors particular to prisons, such as overcrowding, insufficient opportunities to exercise, and the relations between correctional officials and inmates. Whether the predictors of violent behaviour within prisons were actually different from the predictors of violent recidivism in general remained an empirical question.

Until the late 1990s, little work had been done to integrate knowledge about the predictors of institutional violence (Gendreau, Goggin, & Law, 1997). In an unpublished review

paper prepared for the New York Department of Correctional Services in 1981, Chapman and Alexander (as cited in Gendreau et al., 1997) suggested that the inmates' age, intimate partner relationships, and employment status had shown the strongest relationships with prison misconducts. A decade later, Motiuk's (1991) unpublished dissertation indicated that a number of variables from various domains could predict assaults in prison. The best predictors of prison assaults were age (r = -.27) and history of prison misconduct (r = .27). Broadly consistent with the literature on risk for violent recidivism in general, other significant predictors were identified among the domains of education, employment, family/marital status, housing, criminal associates, substance abuse history, and pro-criminal attitudes. Of note, Motiuk found that various criminal history variables including prior convictions and violent records were not significantly related to prison assaults, but they were associated with community follow-up measures such as arrests and general recidivism. Motiuk also found a significant relationship between scores on a structured risk assessment tool, the Level of Service Inventory (LSI; Andrews, 1982), and prison assaults (r = .18).

Gendreau and colleagues (1997) conducted a meta-analysis in an attempt to answer some of the questions that remained regarding predictors of prison misconducts. The authors' chosen outcome measure, prison misconducts, included both violent and non-violent acts, although they noted that the predictor variables produced larger effect sizes for violent outcomes than non-violent outcomes. The results indicated that various personal factors were related to prison misconducts, although antisocial companions (mean weighted r = .28) and prison adjustment (mean weighted r = .32) produced the greatest effect sizes in this domain. Situational factors also proved to be significant predictors of misconduct, with institutional factors (e.g., custody levels, density indexes, and programming) producing a mean weighted r value of .26. Finally, the authors also found that structured risk assessment schemes also predicted prison misconducts, including the LSI-R (mean weighted r = .22) and other risk measures (mean weighted r = .14). Overall, their results suggested that risk for prison misconducts, like violent recidivism, could be meaningfully assessed. Notably, while Gendreau and colleagues observed parallels between the predictors of violent recidivism and those of prison misconducts, they also acknowledged differences.

1.2.4.2 Civil Psychiatric Inpatient Violence. Research on inpatient violence in civil psychiatric settings provides further illustration of contextual factors in the assessment of

violence risk. Unlike correctional settings, admissions to civil psychiatric units do not require individuals to have come into conflict with the criminal justice system, and many authors have argued that the factors that predict violent recidivism in the community may not predict violence within psychiatric settings. For example, Palmstierna and Wistedt (1989) studied aggression among involuntary psychiatric admissions in one Swedish county, and concluded that while patients' histories of violence and substance use were related to inpatient violence, the relationships were weak and of little value to risk assessors.

In search of relevant predictors of aggression in psychiatry, various authors explored an array of personal, clinical, and contextual factors, with varying results. Some researchers identified relationships between inpatient aggression and particular diagnostic categories, including psychotic disorders and mood/bipolar disorders (Flannery, Wyshak, Tecce, & Flannery, 2014; Johnson, 2004; Newton, Elbogen, Brown, Snyder, & Barrick, 2012), while other researchers identified relationships between aggression and particular symptoms, including impulsiveness, hostility, and agitation (Cornaggia, Beghi, Pavone, & Barale, 2011; Newton et al., 2012). While formal criminal history variables per se have not consistently predicted inpatient violence (Cornaggia et al., 2011; Newton et al., 2012), patients' histories of aggression are among the most consistent predictors of civil psychiatric inpatient violence (Amore et al., 2008; Cornaggia et al., 2011; Flannery et al., 2014; Johnson, 2004). Beyond individual predictors, research has also evaluated the utility of structured risk assessment schemes in psychiatric settings. For example, Daffern's (2007) review of a limited number of studies indicated that instruments designed for other settings, including the HCR-20, LSI-R, PCL-R/PCL-SV, VRS, and the VRAG, had shown promise in the long-term prediction of inpatient violence. Daffern also noted however, that some of these instruments may have significant practical limitations when applied to aggression within psychiatric settings, due to factors such as staff training, time constraints, and the importance of acute fluctuations in relevant symptoms/risk factors. With these limitations in mind, Daffern also reviewed data on instruments designed specifically for inpatient settings, including the Brøset Violence Checklist (BVC; see Woods & Almvik, 2002), the Dynamic Appraisal of Situational Aggression: Inpatient Version (DASA: IV; see Ogloff & Daffern, 2006), and the Violence Screening Checklist (VSC; McNiel & Binder, 1994). Again based on limited research, Daffern reported that these brief instruments showed promising predictive validity with regard to short-term risk, and concluded that they were well suited to the

practical requirements of civil psychiatric settings. Subsequent research has further supported the utility and short-term predictive validity of the BVC (Vaaler et al., 2011; Woods, Ashley, Kayto, & Heusdens, 2008) and the DASA (Griffith, Daffern, & Godber, 2013). Thus, while instruments designed for other settings may predict civil psychiatric inpatient violence in the long-term, measures tailored to short-term prediction in civil psychiatric settings appear to be of particular interest to researchers and professionals in this area. This observation is consistent with the suggestion that the selection of risk assessment instruments may be informed by practical considerations, including contexts and intended uses, rather than by predictive validity statistics alone (Douglas, 2014; Yang et al., 2010).

1.2.4.3 A Meta-analysis Combining Correctional and Forensic Psychiatric Samples. Within their landmark meta-analysis comparing the performance of various violence risk assessment instruments, Campbell and colleagues (2009) included institutional violence as a separate outcome variable. Their institutional sample was comprised of roughly equal numbers of effect sizes from general offender and forensic psychiatric samples. Consistent with the findings of Gendreau and colleagues (1997), they found both similarities and differences between the predictors of institutional violence and violent recidivism. In contrast with violent recidivism, institutional violence was better predicted by second-generation instruments, static risk variables, and file-review assessment procedures. Each of the individual tools' performance varied depending on the setting of outcomes. For example, while the HCR-20 predicted institutional violence better than violent recidivism (mean weighted *r*s of .28 and .22), the VRAG had the opposite results (mean weighted *r*s of .15 and .32). Campbell and colleagues' study clearly demonstrated the importance of considering context in risk assessment and the need to validate risk assessment instruments in each setting of interest. With this in mind, the following section will now introduce the setting of the current research.

1.3 Violence and the Forensic Mental Health System

1.3.1 A Unique Population

1.3.1.1 Criminal Responsibility and the NCR Designation. According to Quinsey (2009), the discussion of mental disorders diminishing legal culpability dates back at least to the time of the Roman leader Marcus Aurelius, who argued against conventional criminal punishments for insane offenders over 2000 years ago. The practical implications of this idea have long been a source of uncertainty and as a result, the disposition of mentally ill offenders

has changed over time. For example, in 18th century England it was proposed that an offender should not be held criminally responsible if he lacked the understanding of a 14-year old, or alternatively, if his understanding of his actions was comparable to that of a non-human animal (Quinsey et al., 2006). In a landmark 19th century case, Daniel M'Naghten was acquitted by reason of insanity after he killed a secretary to Britain's Prime Minister (Miller, 2013). M'Naghten was determined to be acting under the delusion that he was the target of a murder plot developed by the Prime Minister and his party. The acquittal was met with public disapproval and led to the formulation of what is known as the M'Naghten test, which states that the offender, due to a disease of the mind, did not comprehend "the nature and quality of the act" (p. 86) or that the act was wrong. The influence of the M'Naghten test is reflected in modern Canadian legislation. Currently in Canada, a defendant may be found Not Criminally Responsible due to a mental disorder (NCR) based on the following criteria found in the *Criminal Code* (Minister of Justice, 2013):

> No person is criminally responsible for an act committed or an omission made while suffering from a mental disorder that rendered the person incapable of appreciating the nature and quality of the act or omission or of knowing that it was wrong. (pp. 37 – 38)

According to the *Criminal Code*, after an NCR finding the court or a provincial Review Board will issue a disposition mandating an absolute discharge, a conditional discharge, or detention in custody in a hospital. Dispositions are made based on the consideration of public safety, the individual's mental state, and the reintegration of the individual back into society.

NCR rulings are rare, they are difficult to procure, and they often result in longer periods of institutionalization than standard sentences for equivalent crimes (Knoll & Resnick, 2008; Reid, 2006). Crocker, Seto, Nicholls, and Cote (2013) recently studied NCR findings in Canada's three most populous provinces and found that over a five year period, the average numbers of NCR findings per year in British Columbia, Ontario, and Quebec, were 393, 144, and 62, respectively. To put these numbers in context, Dauvergne (2012) reported that over a one year period (2010/2011), there were over 400 000 adult criminal court cases in Canada and 64% of those cases resulted in a guilty verdict. According to a research report commissioned by the Canadian Department of Justice (Latimer & Lawrence, 2006) that sampled seven provinces and territories, nearly three quarters of individuals found NCR were mandated to receive psychiatric treatment and more than half were detained in custody in a hospital or equivalent setting.

1.3.1.2 NCR Patients in Hospital. Considering that most NCR patients are detained in hospital settings on specialized forensic units, it is worth considering the characteristics of this population. Currently, the information available regarding NCR individuals in Canada is limited, but the government report produced by Latimer and Lawrence (2006) has provided some useful information regarding these individuals. Demographically, the NCR individuals ranged in age from youth to seniors and were predominantly male (approximately 84%). These researchers found that the most common index offences among the NCR patients were non-sexual assaults (41%) and threats (11%). In all, approximately 77% of the NCR patients had been charged with a non-sexual violent index offence. A further 4% had committed a sexual index offense, while approximately 19% had committed a non-violent index offense. As for the patients' histories of violent behaviour, 34% had a prior conviction for a violent or sexual offense. In contrast, 42% had no previous criminal convictions. With regard to psychiatric diagnoses, the most common for the NCR patients were schizophrenia (52%), substance abuse disorders (31%), affective disorders (28%), and personality disorders (19%). Less common were mental retardation (7%), delusional disorders (6%), and organic brain disorders (3%). Taken together, this information indicates that the NCR population contains a mixture of violent and non-violent individuals, as well as potentially vulnerable individuals, including those with major mental disorders, females, and seniors. It is also important to note that alongside current NCR patients, forensic psychiatric hospitals may also house individuals who are being assessed for criminally responsibility and others for whom the NCR designation may or may not be applicable.

1.3.1.3 Fitness to Stand Trial. In forensic psychiatric facilities, fitness to stand trial is another pertinent legal concept. According to Newby and Faltin (2008), some of the earliest records pertaining to fitness for trial come from English common law in the 17th century. The law required that an individual must have some understanding of the available pleas, the possible outcomes and consequences, and the types of information appropriate for constructing a defense, before being brought to trial. In Canada, fitness for trial transitioned from a somewhat ambiguous legal concept based on case law to a more clearly defined construct in the early 1990s. The relevant legislation from the criminal code (Minister of Justice, 2013) is as follows:

> "unfit to stand trial" means unable on account of mental disorder to conduct a defence at any stage of the proceedings before a verdict is rendered or to instruct counsel to do so,

> and, in particular, unable on account of mental disorder to: (*a*) understand the nature or object of the proceedings, (*b*) understand the possible consequences of the proceedings, or (*c*) communicate with counsel. (p. 14)

Thus, unlike criminal responsibility which pertains to one's mental state at the time of the crime, fitness for trial pertains to one's mental state during the trial (Viljoen, Roesch, Ogloff, & Zapf, 2003). In the event that a judge deems the accused Unfit to Stand Trial (UST), the trial will not proceed. With this in mind, an important feature of fitness for trial is that is considered dynamic; an unfit individual may become fit through treatment, training, or the passage of time, at which point his or her trial may proceed. Individuals may enter the forensic mental health system for fitness-related matters if they are referred for either assessment or treatment services.

1.3.1.4 UST Patients in Hospital. In addition to the discussion of the NCR population, Latimer and Lawrence's (2006) study of mentally disordered accused in the Canadian criminal justice system also addressed the UST population. The most common categories for index offences among the UST accused were non-sexual assaults (45%) and "other non-violent offences" (16%; i.e., non-violent offences excluding arson, break and enter, theft, weapons-related offences). In all, approximately 59% of the UST accused committed a non-sexual violent offence (compared with 77% for the NCR accused). A further 11% had committed sexual offences and 31% committed a non-violent offence. In terms of violent criminal history, 38% of the UST accused had a prior conviction for a violent or sexual offense. In contrast, 44% had no previous criminal convictions. With regard to psychiatric diagnoses, the most common diagnosis for the UST accused was schizophrenia (56%). Rates of other diagnoses were as follows: substance abuse disorders (19%), mental retardation (14%), affective disorders (13%), personality disorders (13%), organic brain disorders (8%), and delusional disorders (4%). Overall, this study indicated that the UST population in Canada represents a group demonstrating diverse psychopathology and with varying histories of violent and criminal behaviour.

1.3.1.5 Combining the At-Risk with the High-Risk. Admission to secure forensic psychiatric facilities is not limited to individuals who have been found NCR or UST. It is worth repeating that these facilities provide assessment services for the criminal justice system, resulting in the admission of a diverse population of accused for whom the NCR and UST designations may or may not ultimately apply. Newby and Faltin (2008) argued that assessments may be requested in the absence of genuine uncertainty regarding the issue at hand (e.g., to delay

a trial), potentially increasing the heterogeneity of the assessment population. Forensic psychiatric facilities also house civilly committed accused or offenders (e.g., Desmarais, Nicholls, Wilson, & Brink, 2012), because they often provide services to acutely ill inmates from correctional facilities. Depending on the setting, forensic mental health settings may also admit individuals for assessments pertaining to Canadian Dangerous Offender (DO) and Long-Term Offender (LTO) designations (Public Safety Canada, 2009); these designations are reserved for high-risk and persistent offenders, and may result in indefinite periods of incarceration, or extended periods of community supervision post-release. Thus, high-security forensic units potentially contain a diverse population including individuals with and without mental illnesses that have come into conflict with the criminal justice system.

As the Latimer and Lawrence (2006) study indicated, many individuals are admitted to these facilities due to criminally violent or aggressive behaviour in the community. Consequently, these units can contain a potentially dangerous mixture of individuals who have exhibited aggressive or predatory behaviour, individuals with gang affiliations, and vulnerable populations, including people with cognitive impairments, people suffering from acute mental illnesses, and women. Physical aggression and victimization are ongoing concerns in these specialized units.

1.3.2 Forensic Psychiatric Inpatient Violence

1.3.2.1 Mental Illness and Violence Risk. Mental disorders are central to the legal issues of criminal responsibility and fitness to stand trial, but their relationship with ongoing violence risk is another matter. The assumption that major mental disorders represent violence risk factors has a long history (Douglas, Guy, & Hart, 2009), but this assumption has been challenged by empirical research. For example, as mentioned previously Bonta and colleagues (1998) did not find a significant relationship between clinical variables and violent outcomes among mentally disordered offenders. Similarly, among the civil psychiatric patients followed in the MacArthur Violence Risk Assessment Study (Rice, Harris, & Quinsey, 2002; Steadman et al., 2000), a diagnosis of schizophrenia was actually negatively associated with violent behaviour in the year following discharge. Other research (Corrigan & Watson, 2005) suggested that psychiatric diagnoses represent significant, if weak, predictors of violence. In the context of these conflicting findings, still further research suggested that the link between mental illness and violence risk is best understood through interactions with other factors. For example, Elbogen

and Johnson's (2009) results from the National Epidemiologic Survey on Alcohol and Related Conditions indicated that major mental illnesses only acted as a significant predictor of violence when they co-occurred with substance abuse or dependence.

Douglas and colleagues (2009) conducted a meta-analysis to further assess the nature of the relationship between psychosis and violence. While their overall findings indicated a small but significant relationship between psychosis and violence, the results also identified moderators of this relationship. For example, the observed relationships between psychosis and violence varied with samples/settings, comparison groups, the presence of other psychopathology, and even with particular symptoms. These findings led Douglas and colleagues to the conclusion that in order for future research on psychosis and violence risk to be informative, it would be important to specify the particular symptoms, settings, and outcomes of interest.

The observation that symptoms of a mental disorder, under certain circumstances, increase an individual's risk for committing violence is not the only important feature of the relationship between mental illness and violence. As Nicholls, Brink, Desmarais, Webster and Martin (2006) noted, persons suffering from mental illness are at an elevated risk of violent victimization. Taken together, the increased risk of violent behaviour and of victimization among persons suffering from major mental disorders makes the study of violence among these populations an important endeavour.

1.3.2.2 Rates and Impacts of Inpatient Violence. Violent incidents on high-security forensic psychiatric units have received less scholarly and practical attention than other related issues, such as violence in the community. It is difficult to quantify the extent of the problem, but existing studies may be used to provide rough estimates of prevalence and incidence rates. In their study of inpatient violence on an Australian forensic psychiatric unit, Daffern, Howells, Ogloff, and Lee (2005) found that 34% of the patients committed an act of violence over a one year period. Nicholls, Brink, Desmarais, Webster, and Martin (2006) found that 39% of their sample of Canadian forensic psychiatric patients committed an act of violence over a one year follow-up period. In another Canadian study, Nicholls, Brink, Greaves, Lussier, and Verdun-Jones (2009) found that of 527 total individuals admitted to a forensic psychiatric inpatient unit, 60% had committed at least one violent act over a one year follow-up period. The average (5.79) and median (1.50) numbers of aggressive incidents per patient indicated that most violent

incidents were perpetrated by a relatively small number of individuals (the most aggressive acts committed by any one person was 161). A study conducted in a Portuguese forensic psychiatric hospital (de Borba Telles, Folino, & Taborda, 2011) found that 74% of patients had committed an act of violence over the course of one year. These estimates, particularly those from Nicholls and colleagues' studies, have important implications for the management of forensic psychiatric inpatient violence in Canada. Firstly, forensic psychiatric inpatient violence appears to be common. Secondly, since not all patients are violent and most incidents are attributable to a relatively small group of patients, it is possible that violence risk assessment practices may be employed to identify and potentially manage a relatively high risk group.

In addition to the obvious consequences of physical injuries to patients, inpatient violence has a variety of other negative impacts which affect various stakeholders. For example, Carmel and Hunter (1989) found that over a one year period at a large forensic state hospital in California, 16% of staff suffered a physical injury due to inpatient aggression. Also reflecting a risk of injury for staff members, Nicholls and colleagues' (2009) results indicated that patients and staff were equally likely to be victims of inpatient aggression. Further research has indicated that inpatient violence negatively affects patients who are not directly victimized and that it damages the therapeutic milieu on psychiatric units (Daffern & Howells, 2002). Beyond the possibility of physical injury, Jacob and Holmes (2011) described some of the psychological consequences experienced by many staff members, including the ongoing experience of fear. Of additional concern to funding agencies and administrators, forensic psychiatric inpatient violence has also been linked with considerable financial costs, related to factors like fluctuating staffing levels (Cyr & Paradis, 2012).

1.3.2.3 Risk Assessment for Forensic Psychiatric Inpatient Violence. Relative to community recidivism and institutional misconducts, risk assessment for forensic psychiatric inpatient violence has received relatively little research attention. As the preceding discussions have indicated, existing risk assessment instruments cannot be assumed to generalize to inpatient violence; even otherwise well supported risk assessment practices should be validated among the particular populations and settings of interest before they are used to inform important decisions. Having reviewed the existing research base, Hogan and Ennis (2010) conducted a meta-analysis of the predictors of violence on secure forensic psychiatric units. The results indicated that the HCR-20 and the Hare Psychopathy scales (having combined effect sizes for the PCL, PCL-R,

and PCL:SV) may have some utility in these settings (mean weighted *r* values of .33 and .26). The Brief Psychiatric Rating Scale (BPRS; Overall & Gorham, 1962), an indicator of mental health symptomatology, produced variable results; unweighted *r* values observed for the BPRS in this review ranged from -0.13 to 0.64. At the time, the number of studies meeting the review's limited inclusion criteria (i.e., forensic psychiatric inpatient sample, violence as part of dependent variable, and a minimum of three separate effect sizes) was quite low. The scarcity of research meant that very few structured instruments could be assessed and that analyses were limited to the main predictive effects of the study measures, averaging across studies. Thus, one of the authors' primary conclusions was that there remained a pressing need for further research and they encouraged the replication of existing studies of risk assessment for forensic inpatient violence.

Since Hogan and Ennis' meta-analysis, a number of additional studies have been published with a focus on structured risk assessment for forensic psychiatric inpatient violence. For example, Chu, Thomas, Ogloff, and Daffern (2011) examined the predictive validity of the Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009) among a sample of 50 forensic psychiatric inpatients. The START was developed specifically for forensic mental health settings, and is comprised of two scales capturing client Vulnerability factors (i.e., risk factors) and Strengths (i.e., protective factors), respectively; the instrument was not examined in Hogan and Ennis' meta-analysis due to a lack of data available at the time. In this study, the dependent variable included incidents of aggression that were observed/documented in the month following the assessments. Using ROC analyses, Chu and colleagues found that START Vulnerability scores and Strength scores were both able to predict inpatient aggression among their sample, as demonstrated by AUC values of .76 and .71, respectively. Using Rice and Harris' (2005) descriptors, these AUC values may be considered large effects.

Desmarais, Nicholls, Wilson, and Brink (2012) conducted an evaluation of the START, the HCR-20, and the PCL:SV among a sample of 120 male forensic inpatients. This archival retrospective study examined the predictive validity of START, HCR-20, and PCL:SV scores for risk over a longer term than Chu and colleagues' (2011) study, with a follow-up time of one year. Results indicated that both START subscales significantly predicted inpatient aggression, as demonstrated by large effects (Vulnerability AUC = .79, Strength AUC = .76). The predictive

validity of the HCR-20 was evaluated using a number of the instrument's risk metrics, all of which significantly predicted inpatient aggression with large effects (total score AUC = .80, SPJ Summary statement = .79, Historical scale AUC = .73, Clinical scale AUC = .74, and Risk Management scale AUC = .77). The PCL:SV also significantly predicted inpatient aggression (AUC = .74). In addition to these primary predictive analyses, Desmarais and colleagues used logistic regression analyses to assess the incremental predictive validity of START scores beyond the static factors of the HCR-20 Historical scale and PCL:SV. START Vulnerability scores were found to incrementally predict inpatient aggression beyond HCR-20 Historical scale scores and PCL:SV scores, but START Strength scores were not.

Vitacco, Gonsalves, Tomony, Smith, and Lishner (2012b) studied the predictive validity of structured risk instruments in a sample of 103 male forensic inpatients who had been found Not Guilty by Reason of Insanity (NGRI). These authors collected observations from a follow-up period of six months to capture incidents of inpatient violence. Measures included: Factor 1 and Factor 2 of the PCL-R, the Historical, Clinical, and Risk Management scales of the HCR-20, and the total score of the VRAG. Using *t*-tests to compare the scores of aggressive and non-aggressive patient groups, the authors observed significant differences on each of these measures, with the exception of the HCR-20 Risk Management scale. Logistic regression analyses were also conducted to assess the incremental predictive validity of dynamic factors, as measured by the HCR-20 Clinical and Risk Management scales, beyond the static factors of the HCR-20 Historical scale, VRAG, and PCL:SV. While a model containing only dynamic predictors was found to significantly predict inpatient violence, the dynamic predictors did not demonstrate unique and independent prediction beyond the static measures. The authors also noted that a similar pattern of results was observed for static measures, in that they significantly predicted inpatient aggression on their own, but did not demonstrate incremental prediction beyond the dynamic measures.

Wilson, Desmarais, Nicholls, Hart, and Brink (2013) analyzed data from 30 male forensic inpatients, to assess the predictive validity of the HCR-20 and the START for inpatient aggression. In addition to their goal of assessing the predictive validity of the measures for inpatient aggression, and consistent with the discussion of dynamic risk factors presented earlier in the current paper, Wilson and colleagues also designed this research to evaluate whether purportedly dynamic risk factors would perform as intended. Follow-up was divided into four

distinct three month periods, and participants were assessed prior to every period; rather than individual patients, each assessment and corresponding follow-up period was treated as a unit of analysis. Given that dynamic scores for individual patients varied across the assessment periods, the investigators suggested that if dynamic scores were found to predict aggression overall, then it could be inferred that the risk factors were genuinely dynamic predictors. To examine the first research question, they evaluated various elements of the HCR-20 using ROC analyses, including: each individual scale, a combination of the Clinical and Risk Management scales, the total score, and the SPJ summary judgement. Each HCR-20 risk metric significantly predicted inpatient aggression averaging across the four follow-up periods, with large AUC values ranging from .79 to .91. The START was evaluated in the same way, using Vulnerability scores, Strength scores, and the SPJ summary risk judgement; each of these measures produced large AUC values, ranging from .82 to .89. Survival analyses were then used to assess the incremental validity of the dynamic factors of the HCR-20, as measured by summing the Clinical and Risk Management scores, beyond the static factors of the Historical scale. The results indicated that the dynamic factors did demonstrate incremental prediction beyond the Historical scale. An additional survival analysis was employed to assess the incremental validity of the START scores beyond static risk factors, again measured by the HCR-20 Historical scale. START Vulnerability scores, but not Strength scores, demonstrated incremental predictive validity beyond the HCR-20 Historical scale.

The studies described above have focused on risk for aggression over follow-up periods of one month or more, but researchers have also considered more acute risk variables, which might be applicable over periods of hours or days. Chu, Daffern, and Ogloff (2013) for example, assessed the predictive validity of three instruments, the Brøset Violence Checklist (BVC; see Woods & Almvik, 2002), the Dynamic Appraisal of Situational Aggression (DASA; Ogloff & Daffern, 2002), and the HCR-20 Clinical scale, for the prediction of inpatient aggression over 24 hour follow-up periods. Daily scores produced by these measures, which were all designed to capture acute/clinical risk factors (e.g., impulsivity, irritability, and acute symptoms of mental illness), significantly predicted inpatient aggression over the short-term (BVC AUC = .77, DASA AUC = .76, and HCR-20 Clinical scale AUC = .68). A recent study conducted by Woods, Olver, and Mueller (2015) evaluated the predictive validity of the BVC for inpatient aggression occurring over follow-up periods equal to the length of single nursing shifts (there

were two shifts per day). Among this sample of Canadian forensic psychiatric inpatients, the BVC demonstrated strong predictive validity, with an AUC of .73.

Recently, further published studies have presented psychometric data pertaining to specialized tools developed to address local needs for structured risk assessment for inpatient violence. Chagigiorgis, Michel, Seto, Laprade, and Ahmed (2013) described results of an evaluation of the Brockville Risk Checklist (BRC; Ahmed, Adamson, & Laprade, 2004), which was designed to structure institutional risk assessments conducted at regular case conferences. Items, or risk factors, on the BRC are divided among four overlapping risk scales, entitled Harm to Others (e.g., control-override symptoms), Harm to Self (e.g., actual or threatened self-harm), Risk of Neglect (e.g., inadequate attention to nutritional needs), and Risk of Exploitation by Others (e.g., indiscriminate giving away of personal effects). Chagigiorgis and colleagues evaluated the predictive validity of the instrument among a sample of 121 forensic psychiatric inpatients. The investigators reported that higher scores on the Harm to Others scale were associated with increased risk of aggression during the corresponding follow-up periods of approximately four to six weeks. Elsewhere, Starzomski and Wilson (2015) evaluated an instrument called the Imminent Risk Rating Scale (IRRS), which was intended to assess risk over shorter follow-up periods. According to Starzomski and Wilson, this instrument was developed in response to the perceived limitations of various existing violence risk assessment instruments, as applied to the assessment forensic psychiatric inpatient violence. They suggested that tools like the HCR-20 and VRAG were not optimal because they were not sensitive to short-term fluctuations in risk, while other tools designed to capture acute risk factors, like the BVC and DASA, were insensitive to important contextual factors, such as hospital milieu. Thus, the IRRS was developed as a screening tool to capture seven risk factors, such as criminal history and problems with the context/environment (e.g., crowding). Starzomski and Wilson evaluated the predictive validity of the measure among a total of 121 forensic inpatients across four assessment periods and found that scores predicted aggression over corresponding follow-up periods of less than one week (AUCs of .69 to .74).

Overall, these results indicate that there remains a need for further study of risk assessment for forensic psychiatric inpatient violence. Indeed, the development of instruments like the BRC and IRRS reflects the respective developers' impressions of a continued need for adequate measures of risk in these settings. Other researchers have evaluated the predictive

validity of other tools, such as the START and HCR-20, but the limited number of studies has left a number of questions to be answered about the efficacy and effectiveness of existing risk assessment instruments. For example, forensic mental health settings represent, in many ways, an intersection between psychiatric and criminal justice settings, but the extent to which risk factors for institutional violence validated in these other settings are each applicable to violence in forensic hospitals is still uncertain. Thus, it is left to clinicians and researchers alike to continue to ask: which violence risk assessment instruments are valid for the assessment of risk for inpatient violence? What are the relative contributions of acute/clinical risk variables and more stable risk factors to the assessment of risk for inpatient aggression? What information is required to conduct a valid assessment of inpatient violence risk (e.g., file information only versus interview, plus file review)? And for how long may a given assessment be considered valid?

The previous research has indicated, at least, that valid risk assessment in forensic psychiatric settings is possible. Certainly, acute/clinical risk factors and more stable risk factors each appear to be worth investigating further, given that each type of risk factor has unique implications for intervention and risk management activities (i.e., training for crisis intervention/avoidance versus long-term formal risk reduction interventions). With the long-term aim of decreasing the serious negative consequences of inpatient aggression in mind, establishing the fundamental predictive validity of existing and emerging instruments for this outcome remains an important goal.

1.4 Purpose of the Proposed Research

The current program of research was designed to assess the validity of existing violence risk assessment instruments on secure forensic psychiatric units, with a focus on the following four research objectives: 1) observing the profile of dynamic changes in violence risk detected by existing violence risk assessment instruments, namely the VRS, the START, and the HCR-20^{V3}, during admissions to forensic psychiatric units; 2) assessing whether existing violence risk assessment instruments, namely the HCR-20^{V3}, the PCL-R, the START, the VRS, and the VRAG-R, can be used to assess risk for inpatient violence on forensic psychiatric inpatient units; 3) evaluating the contribution of dynamic risk measures to the prediction of inpatient violence on high security forensic psychiatric units; and 4) assessing the relationship between dynamic changes in risk and inpatient violence. These research objectives were addressed through two

studies. Study 1 was archival, exploratory, descriptive, and retrospective, while Study 2 was prospective. Another explicit aim of this program of research was to produce results with practical applications in forensic psychiatric hospitals. Thus, the design and analyses were considered from the perspective of professionals looking for useful and feasible methods for assessing risk for inpatient violence on forensic psychiatric units.

1.4.1 Research Objective 1: Detecting Dynamic Risk: The VRS, the START, and the HCR-20^{V3}

Existing dynamic violence risk assessment instruments have not yet been proven to perform in the assessment of risk for inpatient violence in secure forensic psychiatric hospitals. That is, dynamic violence risk factors that have been identified and validated in other settings may be less applicable to assessments of risk for inpatient violence. As Olver and Wong (2011) argued, for a dynamic risk factor to be reliably linked to violent outcomes, it must be the product of a "reliable and credible change agent" (p. 123). An example of such a change agent is an empirically supported treatment program, with associated measurable indicators of success. With this in mind, it follows that the correspondence between a given risk instrument and the relevant change agents in a particular setting will be a determinant of predictive success. The VRS for example, is based on a stages-of-change-model (Wong & Gordon, 2006) and is well-suited to capturing changes that result from formal violence-reduction treatment programming, but patients on secure forensic psychiatric hospital units vary on the extent to which they receive this type of programming. For example, a particular patient may primarily be treated with psychiatric medication to stabilize his or her mental state and to facilitate discharge back to a correctional facility or transfer to a lower security unit. As for the HCR-20^{V3}, it contains certain items, such as active symptoms of a mental illness, that appear to be well suited to detecting changes during even brief admissions to these units, while others, such as negative attitudes, may be less so. Ostensibly, the START, with its greater focus on mental health symptoms, is perhaps better suited to capturing changes on forensic psychiatric units overall, relative to the VRS and the HCR-20^{V3}. As a first step towards evaluating the study instruments' relevance to dynamic risk assessment on secure forensic psychiatric units, the first research objective involved assessing the presence and size of risk changes detected by each dynamic measure over the course of admissions.

1.4.2 Research Objective 2: Assessment of Risk for Inpatient Violence on Forensic Psychiatric Inpatient Units

The second research objective was to establish the predictive validity of existing structured instruments for forensic psychiatric inpatient violence. Thus, the relationships between scores on each of the structured instruments, including their respective subscales, and inpatient violence were evaluated in two separate studies.

The following instruments were investigated: the HCR-20^{V3}, the PCL-R, the START, the VRS, and the VRAG-R. These instruments were selected for the following reasons: to build upon the evidence provided by the current author's previous meta-analytic study (i.e., the HCR-20 scheme and PCL-R; Hogan & Ennis, 2010) and to investigate other existing static (i.e., the VRAG-R) and dynamic violence risk instruments (i.e., VRS and START).

1.4.3 Research Objective 3: Contribution of Dynamic Risk Measures to the Prediction of Inpatient Violence

As was previously discussed, it has become increasingly accepted among forensic professionals that violence risk is dynamic and so assessment should incorporate dynamic or changeable risk factors. The current research investigated the incremental validity of dynamic risk factors by assessing whether they provided increased accuracy in forecasting violence risk on forensic psychiatric units, beyond historical/static measures.

1.4.4 Research Objective 4: The Relationship between Dynamic Changes in Risk and Inpatient Violence

The current research also evaluated the relationship between dynamic changes in risk and inpatient violence. For a risk factor to be considered truly dynamic, it should be associated with observable changes, and those changes should be related to changes in risk (Andrews & Bonta, 2010; Douglas & Skeem, 2005; Kraemer et al., 1997). While the predictive validity of change scores for subsequent risk could not be evaluated in the current research, the concurrent validity of dynamic changes in risk and inpatient aggression was evaluated through Study 1.

Chapter 2: STUDY 1

2.1 Introduction

This archival study was designed to evaluate whether a number of existing forensic instruments were applicable to the assessment of risk for forensic psychiatric inpatient violence. Beyond the primary evaluations of the predictive validity of individual tools for inpatient violence, further investigations were conducted to assess the relative contributions of different types of risk factors to the assessment of risk for inpatient outcomes. Additionally, this study was designed to explore the utility of a selection of these measures for the assessment of dynamic changes in risk among forensic psychiatric inpatients, as well as the relationship between measurable dynamic change and concurrent inpatient violence. Evaluating the predictive validity of dynamic change scores involves the observation of outcomes subsequent to the measurement of change (Andrews & Bonta, 2010; Douglas & Skeem, 2005; Kraemer et al., 1997), and so it should be noted that the change analyses contained herein are limited to concurrent validity. Due to the exploratory nature of the initial questions regarding dynamic risk, one goal was simply to provide a descriptive account of the profile of change scores produced by patients on these units between admission and discharge/transfer. The following hypotheses and research questions were addressed.

2.1.1 Hypotheses/Research Questions

2.1.1.1 Research Objective 1: Detecting Dynamic Risk: The VRS, START, and HCR-20^{V3}.

1.1. Change scores, produced by computing the differences between scores at admission and at discharge from the hospital, were predicted to be significantly greater than zero, for the following instruments: HCR-20^{V3} Clinical Scale, HCR-20^{V3} total score, START total score.

1.2. It was expected that change scores produced by the VRS Total Score would not be significantly greater than zero, due to the variability in formal violence reduction programming provided to participants.

2.1.1.2 Research Objective 2: Prediction of Inpatient Violence on Forensic Psychiatric Inpatient Units.

2.1. It was anticipated that higher pretreatment risk scores, computed from the HCR-20^{V3}, PCL-R, START, VRS, and VRAG-R, would be associated with a greater probability of a violent incident within the institution over the follow-up period.

2.2. Using Receiver Operating Characteristic (ROC) analyses, it was expected that HCR-20^{V3}, PCL-R, START, VRS, and VRAG-R pretreatment scores would demonstrate predictive accuracy (i.e., AUC values significantly greater than .50) with respect to outcome measures of inpatient violence.

2.1.1.3 Research Objective 3: Contribution of Dynamic Risk Measures to the Prediction of Inpatient Violence.

3.1. It was expected that dynamic factors of the VRS and HCR-20^{V3} Clinical Scale would demonstrate incremental validity for the prediction of institutional violence, after controlling for the static factors of the VRS and HCR-20^{V3} Historical Scale, respectively.

3.2. Dynamic scores, produced by the VRS, START, and the HCR-20^{V3} Clinical scale, were entered into logistic regression analyses along with a measure of static risk, the VRAG-R, to determine the unique and independent contributions of each to the prediction of the dichotomous outcome variable (any violent incident in the follow-up period).

2.1.1.4 Research Objective 4: The Relationship between Dynamic Changes in Risk and Inpatient Violence.

4.1. It was hypothesized that greater change scores, produced by the HCR-20^{V3}, START, and VRS, would be associated with a decrease in the probability of having engaged in inpatient violence during the index admission.

2.2 Method

2.2.1 Participants

The participants in Study 1 included a sample of adult forensic psychiatric inpatients, admitted to a maximum security unit in the Acute Assessment and Treatment Program, Forensic Psychiatry Services, Alberta Hospital Edmonton (AHE). A summary of key descriptive statistics and demographic variables is presented in Table 2.1. The sample was comprised of 99 consecutively admitted patients, with a length of stay of at least two weeks, and with admission dates ranging from 2005 to 2008. Some participant files contained sufficient data to code particular study instruments, but not outcome data. The author chose to include these individuals in the overall sample and to use their data in non-predictive analyses, such as convergent validity analyses, because these analyses were deemed to be of significant value, particularly for the newer study instruments (i.e., HCR-20^{V3} and VRAG-R). Mean age was 36.7 years (*SD* = 12.8) at admission and 38.3 years (*SD* = 12.3) at discharge. The majority of the sample was male

Table 2.1

Summary of Relevant Demographic, Mental Health, and Historical Variables

	Mean (SD)	%
Demographics		
Age at Admission (years)	36.7 (12.8)	
Age at Discharge (years)	38.3 (12.3)	
Male		86
Female		14
White		61
Aboriginal Descent		12
Education		
Unknown		24
Less than Grade 12		37
High School Diploma/GED		15
Any Post-Secondary		23
Admission Status		
NCR		52
Certified Prisoner		24
DO/LTO Assessment		10
Unfit		9
Other		6
Mental Health		
Schizophrenia/Psychotic Disorder		61
Antisocial Personality Disorder		22
Substance/Alcohol Diagnosis		70
Historical		
Any Previous Criminal Charge		68
Violent Index		71
Previous Institutional Aggression		54

Note. $n = 99$.

(86%) and 14% of the sample was female. Approximately 61% of the sample was identified as white, 12% was identified as being of Canadian Aboriginal descent, and the remaining participants were identified as being of diverse ethnic backgrounds. Information regarding educational achievement was unavailable for approximately one quarter of the sample (24%); because these individuals represented such a large proportion of the sample, they were included as a separate category (i.e., "unknown") in the calculation of percentages. According to the available documentation, approximately 15% of the sample had obtained a high school diploma or equivalent, while approximately one third of the sample (33%) had completed at least some post-secondary training, through a college, university, or professional school. Just over half of the sample (52%) had been found NCR, approximately one quarter (24%) were convicted/remanded prisoners admitted for mental health services, and the remaining one quarter were admitted for a variety of other forensic assessment and treatment services. Over two thirds of the sample had at least one previous criminal charge (68%) and over half (54%) had a documented history of institutional aggression, in either a hospital or prison. The majority of the sample (71%) was admitted for violent index offences.

2.2.2 Program Description

The Acute Assessment and Treatment Program at AHE conducts remand assessments and provides psychiatric treatment services for individuals who have come into contact with the criminal justice system. The program's catchment area includes Edmonton, northern Alberta, the Northwest Territories, the Yukon, and a portion of Nunavut. Services are provided by a multi-disciplinary team, including nursing staff, medical doctors, psychologists, social workers, occupational therapists, and recreational therapists. Treatment services are provided primarily to individuals who have been deemed NCR, but are also provided to other individuals, including those who have been deemed unfit to stand trial and prisoners who have been certified under the Mental Health Act. Patients may receive a variety of biological, social, and psychological interventions depending on individual needs. In addition to individual counselling, a variety of psychoeducational groups have traditionally been offered on a rotating basis, addressing topics such as mental health, substance abuse, anger management, relapse prevention, and social skills. Depending on legal constraints, patients may also participate in recreational and occupational programming within and outside the hospital, to facilitate safe reintegration back to the community.

2.2.3 Materials

The study materials consisted of: the risk/forensic assessment instruments, including the HCR-20^{V3}, the PCL-R, the START, the VRAG-R, and the VRS; an inpatient violence outcome measure, the Staff Observation Aggression Scale-Revised (SOAS-R; Nijman et al., 1999), which was used to provide an operational definition of inpatient aggression; and an additional data collection protocol developed for the study.

2.2.3.1 HCR-20 Version 3. See Appendix A. The recently published Historical-Clinical-Risk Management-20 Version 3 (HCR-20^{V3}; Douglas, Hart, Webster, & Belfrage, 2011) is the latest iteration of the widely used SPJ violence risk assessment scheme. Its predecessor, the HCR-20 Version 2 (Webster, Douglas, Eaves, & Hart, 1997) has been widely accepted as a useful tool in violence risk assessment, by both forensic professionals and by the courts. It was identified as one of the most commonly used tools in violence risk assessments among forensic psychologists (Archer et al., 2006) and was recently found to be regularly used and accepted in legal settings in the United States (Vitacco, Erickson, Kurus, & Apple, 2012). Additionally, a large body of research supports the predictive validity of the HCR-20 risk assessment scheme (Version 1: Webster, Eaves, Douglas, & Wintrup, 1995; Version 2: Webster, Douglas, Eaves, & Hart, 1997) across a wide range of settings. For example, researchers have found that the HCR-20 has provided valid and useful assessments of violence risk among civil psychiatric populations (see Douglas, Ogloff, Nicholls, & Grant, 1999; McNiel, Gregory, Lam, Binder, & Sullivan, 2003; Nicholls, Ogloff, & Douglas, 2004) and convicted offender populations (see Douglas & Webster, 1999; Neves, Goncalves, & Palma-Oliveira, 2011). According to Douglas (2014), early results using the HCR-20 Version 3 have also supported its predictive validity in forensic contexts.

The HCR-20 scheme is used to guide professional judgement and to assist assessors in identifying factors that may influence one's risk for violence (Guy, Douglas, & Hendry, 2010). This instrument includes twenty items separated into three sections: historical risk factors (e.g., history of violence), clinical factors that are relevant to current risk (e.g., mental health status), and risk management factors that may influence future risk (e.g., inadequate social support). Each item is scored on a three-point scale (not present, possibly or partially present, or present). On the balance of information, clinicians arrive at an overall risk decision of low, moderate, or high, with regard to Risk for Future Violence or Case Prioritization. Using the updated HCR-20

Version 3, assessors may also arrive at two further overall risk decisions: Risk for Serious Physical Harm and Risk for Imminent Violence. Researchers have commonly computed a total score out of 40 to assess predictive validity of the HCR-20 scheme (Douglas et al., 2014) and this method was employed in the current investigation. However, given that the newest iteration of the HCR-20 scheme involves separate scores for the presence and relevance of each risk factor, two separate total scores were computed.

2.2.3.2 PCL-R. See Appendix B. The Psychopathy Checklist Revised Version (PCL-R; Hare, 1991) was designed to measure the personality construct of psychopathy, which taps a number of affective, interpersonal, and behavioural characteristics (Hare & Neumann, 2009). Examples of the characteristics assessed by the PCL-R are a lack of empathy, impulsivity, and antisocial behaviour. The 20 PCL-R items are scored on a three-point scale (0, 1, or 2) and may be summed to compute a total score. Additionally, research has identified two primary conceptualizations of the factor structure of the PCL-R (Hare & Neumann, 2008): a four facet model, comprised of Interpersonal (Facet 1), Affective (Facet 2), Lifestyle (Facet 3), and Antisocial (Facet 4) facets; and a two factor model, with items divided more broadly into the Interpersonal/Affective (Factor 1) and Social Deviance (Factor 2) components.

The PCL-R and its derivatives were found to be the most commonly used tools for Risk Assessment/Psychopathy assessments among forensic psychologists (Archer et al., 2006). In their meta-analysis, Campbell, French, and Gendreau (2009) found that the Hare psychopathy scales were associated with institutional violence and violent recidivism among adult offenders. The Hare Psychopathy scales have also received research support for the prediction of forensic psychiatric inpatient violence (Hogan & Ennis, 2010).

2.2.3.3 START. See Appendix C. The Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009) is a clinical guide for the assessment of dynamic risk factors as well as strengths and treatment-related variables. Originally constructed for use in forensic psychiatric hospitals (Doyle, Lewis, & Brisbane, 2008), the START has shown promise for the prediction of violence in varied settings. Two studies of mixed samples of civil and forensic psychiatric inpatients found modest relationships between START scores and violence (Gray et al., 2011) and violence against staff (Nonstad et al., 2010) respectively. Braithwaite, Charette, Crocker, and Reyes (2010) found a modest relationship between START scores and aggression towards others in a Canadian civil

psychiatric setting. In a study of inpatient aggression on a forensic psychiatric unit, Chu and colleagues (2011) found an acceptable relationship between START Risk scores and interpersonal violence. Desmarais and colleagues (2012) also found support for the START in the prediction of inpatient aggression. While the START has less research support than more established tools like the HCR-20, as Chu and colleagues observed, existing evidence has supported the utility of the START for the prediction of forensic psychiatric inpatient violence. The START includes a total of 20 factors that may be scored on a three-point scale to compute both a Strength total score and a Vulnerability total score (Desmarais et al., 2012). For the current investigation, both Vulnerability (i.e., risk) and Strength scores were computed; due to resource and time considerations, the SPJ Summary statements associated with the START were not assessed.

2.2.3.4 VRAG-R. See Appendix D. As described previously, the Violence Risk Appraisal Guide (VRAG; Quinsey, Harris, Rice, & Cormier, 1998; 2006) is an empirically derived risk assessment instrument which has been found to predict community and institutional violence (e.g., Campbell et al., 2009). Recently, Rice, Harris, and Lang (2013) published an article providing support for an updated version of the tool, the Violence Risk Appraisal Guide – Revised (VRAG-R). This article suggested that the revised version had improved upon the original instrument in terms of both predictive validity and useability. As a result, this most recent version of the instrument was employed in the current investigation, to assess its predictive validity for inpatient violence. The instrument includes 12 items, which may add or subtract varying amounts from the total score. Examples of VRAG-R items include marital status, criminal history, and antisociality.

2.2.3.5 VRS. See Appendix E. The Violence Risk Scale (VRS; Wong & Gordon, 1999-2003) is a specialized tool designed to allow professionals to combine risk assessment and treatment considerations (Wong & Gordon, 2006). The VRS is based on the Risk-Need-Responsivity model (Andrews et al., 1990) and informs an evaluation of risk for violence, dynamic risk factors which may be targeted for treatment, an offender's readiness for treatment, and positive treatment-related changes (Wong & Gordon, 2006). In their meta-analysis, Yang, Wong and Coid (2010) found that the VRS predicted violence with a moderate level of accuracy. Dolan and colleagues (Dolan & Fullam, 2007; Dolan, Fullam, Logan, & Davies, 2008) have also provided preliminary evidence indicating that the VRS may be useful for predicting forensic

psychiatric inpatient violence. The VRS includes 26 items scored on a four-point scale (0, 1, 2, or 3), which may be combined to produce a total score. Additionally, VRS scores may be broken down into two subtotals, pertaining to static items and the dynamic items, respectively.

2.2.3.6 SOAS-R. See Appendix F. The Staff Observation Aggression Scale – Revised (SOAS-R; Nijman et al., 1999) is an empirically supported (Nijman, Palmstierna, Almvik, & Stolker, 2005) instrument developed for documenting aggressive incidents. The SOAS-R was used to identify an operational definition of inpatient aggression for the current study. Aggression is defined as, "any verbal, nonverbal, or physical behavior that was threatening (to self, others, or property), or physical behavior that actually did harm (to self, others, or property)" (p. 208). The SOAS-R allows users to document the circumstances provoking aggression, as well as the nature, severity, and consequences of each incident; these functions were not utilized in the current study, due to raw data limitations.

2.2.3.7 Operational Definition of Inpatient Violence. Consistent with the chosen outcome measure, the SOAS-R, outcomes were defined as "any verbal, nonverbal, or physical behavior that was threatening (to self, others, or property), or physical behavior that actually did harm (to self, others, or property)" (Morrison, 1990). It should be noted that this operational definition is relatively broad and it includes behaviours that would not necessarily result in physical injury to a specific individual; thus, this definition has the potential to increase the base rate of the outcome variable. The decision to use such a definition was taken because evidence has suggested that incidents of inpatient violence tend to be underreported (Woods et al., 2015) and to avoid excluding serious incidents based on their consequences alone (e.g., one patient's attempt to strike another individual with an object, that was narrowly prevented by nursing staff). Nonetheless, caution was taken to limit outcomes to serious incidents (such as the example provided above). For reference, the terms *inpatient violence* and *inpatient aggression* will be used interchangeably in this document. Violent incidents were coded as a binary outcome variable.

2.2.3.8 Data Collection Protocol. See Appendix G. A data collection protocol was drafted to gather data pertaining to the current program of research. The protocol contains demographic information, as well as information pertaining to the risk instruments, victim characteristics, and psychiatric information.

2.2.4 Procedure

Ethical approval for this research was obtained from the Behavioural Research Ethics Board of the University of Saskatchewan,, as well as the Health Research Ethics Board – Health Panel of the University of Alberta. These research ethics boards (REBs) provided ethical review and approval of the investigator's proposed methods, including the use of institutional records to score risk instruments. Additional formal administrative approvals were obtained from Alberta Health Services and Alberta Hospital Edmonton, which permitted the writer to access the site and to review archival patient files. It is also noted that given the archival nature of the study, consent was not obtained directly from individual patients. In the province of Alberta, the University of Alberta's REB possessed the authority to provide consent for researchers to access this type of data based on the *Health Information Act* (Government of Alberta, 2000):

> (i) the proposed research is of sufficient importance that the public interest in the proposed research outweighs to a substantial degree the public interest in protecting the privacy of the individuals who are the subjects of the health information to be used in the research, (ii) the researcher is qualified to carry out the research, (iii) adequate safeguards will be in place at the time the research will be carried out to protect the privacy of the individuals who are the subjects of the health information to be used in the research and the confidentiality of that information, and (iv) obtaining the consents referred to in clause (a) is unreasonable, impractical or not feasible (pp. 36-37)

All study instruments were scored based on archival information obtained from multidisciplinary institutional records, by the author and one trained research assistant, both graduate students with clinical and research experience related to risk assessment. Both coders also received training on all study instruments from the author's doctoral research supervisor, a Registered Doctoral Psychologist in the Province of Saskatchewan, with experience conducting risk assessments. To keep the coders blind to outcomes and avoid criterion contamination, clinical records at *pretreatment* were separated from the rest of the files to facilitate coding. Records dated no later than the patient's first mandatory report prepared for the Alberta Board of Review (45 days after admission or sooner) were included as the pretreatment documents. These records were used to code all of the study instruments. Once pretreatment scoring was complete, the coders used the entire clinical file to compute a *posttreatment* score on the HCR-20^{V3}, VRS, and START, using all information available at the date of discharge from the hospital, or December 31st, 2013 if the patient was never discharged. Records and information from the final

30 days of admission were emphasized in the ratings of clinical/dynamic items. After instrument scoring was complete, the coders scanned the files for any evidence of inpatient aggression during the index admission. Because exact dates could not be identified for all incidents of inpatient aggression, outcome variables and analyses requiring time data (e.g., survival analyses) were not used for institutional outcomes.

2.2.5 Data Analytic Plan

2.2.5.1 Preliminary Analyses: Inter-rater Reliability and Convergent Validity. The interrater reliability of the instruments was assessed by computing single measure intraclass correlation coefficients (ICCs). Pearson correlations were also computed among the measures, thereby examining their convergent validity.

2.2.5.2 Research Objective 1: Detecting Dynamic Risk: The VRS, START, and HCR-20^{V3}. A number of *t*-tests were conducted to examine whether each dynamic instrument detected measurable changes between admission and discharge/transfer (i.e., to another unit).

2.2.5.3 Research Objective 2: Prediction of Inpatient Violence on Forensic Psychiatric Inpatient Units. Due to the absence of adequate time data, analyses of predictive validity for inpatient violence were limited to those procedures that do not require time-based variables. ROC analyses producing area-under-the-curve (AUC) statistics were conducted on the HCR-20^{V3}, PCL-R, START, VRS, and VRAG-R pretreatment scores, with respect to inpatient violence. ROC analyses provide an indication of the probability that a randomly selected recidivist would score higher on the predictor, in this case a risk measure, than a randomly selected non-recidivist (Andrews & Bonta, 2010). An AUC of .50 would indicate that the predictor performs no better than chance, while an AUC of 1.00 would indicate that the predictor is able to perfectly discriminate between recidivists and non-recidivists. In addition to total scores and component scores of the various measures, HCR-20^{V3} SPJ summary statements were treated and analyzed as numerical scores (Desmarais et al., 2012; Gray et al., 2011; Wilson et al., 2013). Scores on the risk instruments were coded as continuous variables, while inpatient violence was coded as a dichotomous outcome variable (any versus no incidents of inpatient violence during the index admission).

2.2.5.4 Research Objective 3: Contribution of Dynamic Risk Measures to the Prediction of Inpatient Violence. Using the dichotomous inpatient violence outcome variable, hierarchical logistic regression analyses were conducted to determine whether the dynamic

scales demonstrated incremental validity over their corresponding historical/static scales. Logistic regression analyses were also used to assess the unique and independent contributions of the various dynamic scores compared with the VRAG-R. Logistic regression analyses produce exponentiated beta coefficients (e^B), which represent odds ratios reflecting the change in odds of the dependent variable occurring, that is associated with a one point increase in the independent variable (Tabachnick & Fidell, 2013). For example, an e^B of 1.10 for a given risk measure in these analyses would indicate that an increase of one point was associated with a 10% increase in the odds of inpatient violence.

2.2.5.5 Research Objective 4: The Relationship between Dynamic Changes in Risk and Inpatient Violence. Hierarchical logistic regression analyses were also conducted to investigate whether the dynamic change scores produced by the VRS, START, and HCR-20^{V3} would demonstrate incremental concurrent validity with inpatient violence, controlling for baseline risk, using the dichotomous outcome variable.

2.3 Results

It should be noted that limitations in the data available in institutional files influenced the analyses and conclusions herein. For example, the majority of the institutional files reviewed for this study contained sufficient information to code some, but not all the of the study measures. In order to make use of as much of the available data as possible, and to maximize statistical power for the various analyses, every participant with a score for given instrument or subscale was included in individual analyses. As a result, sample sizes vary by instrument and by analysis. Overall descriptive statistics, including the sample size, mean scores and standard deviations pertaining to each instrument are presented in Table 2.2.

Overall, with the exception of the PCL-R scores, the distributions of scores the risk instruments were comparable to similar studies. More specifically: HCR-20^{V3} subscale scores were similar to the values reported in the most recent HCR-20^{V3} research bibliography, taken from an offender sample (Douglas et al., 2014); START risk scores were similar to the scores reported in Desmarais and colleagues' (2012) and Wilson and colleagues' (2013) studies of forensic inpatients; the mean VRAG-R score fell between the 54th and 55th percentiles reported in the scoring manual; and the mean VRS total score was a few points below the mean reported in Wong and Gordon's (2006) normative study of 918 offenders. The mean PCL-R score on the

Table 2.2

Descriptive Statistics: Means and Standard Deviations of Pretreatment and Posttreatment Scores

Instrument	*n*	Mean	SD
HCR-20 H Pres. (PreTx)	90	12.61	4.00
HCR-20 H Pres. (PostTx)	82	12.48	4.25
HCR-20 C Pres. (PreTx)	90	5.23	2.93
HCR-20 C Pres. (PostTx)	82	3.02	2.32
HCR-20 RM Pres. (PreTx)	89	7.07	1.92
HCR-20 RM Pres. (PostTx)	82	6.23	2.25
HCR-20 Total Pres. (PreTx)	89	24.89	7.19
HCR-20 Total Pres. (PostTx)	82	21.99	7.07
HCR-20 H Rel. (PreTx)	89	10.96	4.32
HCR-20 H Rel. (PostTx)	81	10.99	4.27
HCR-20 C Rel. (PreTx)	90	6.17	2.85
HCR-20 C Rel. (PostTx)	82	5.26	2.69
HCR-20 RM Rel. (PreTx)	89	7.19	2.05
HCR-20 RM Rel. (PostTx)	82	6.68	2.37
HCR-20 Total Rel. (PostTx)	88	23.93	7.98
HCR-20 Total Rel. (PreTx)	81	22.54	8.36
START Vuln. (PreTx)	91	19.45	9.11
START Vuln. (PostTx)	81	12.70	7.56
START Str. (PreTx)	90	4.83	6.90
START Str. (PostTx)	81	8.63	8.33
VRS Static (PreTx)	77	7.30	4.39
VRS Static (PostTx)	74	7.08	4.35
VRS Dynamic (PreTx)	76	30.86	11.16
VRS Dynamic (PostTx)	73	30.15	11.52
VRS Total (PreTx)	77	38.14	14.65
VRS Total (PostTx)	73	37.02	14.81
VRAG-R	76	1.24	20.15
PCL-R	77	12.87	8.99

Note. PreTx = Pretreatment score. PostTx = Posttreatmetn score.
For the HCR-20, H = Historical scale, C = Clinical scale, R = Risk Management scale, Pres. = Presence score, Rel. = Relevance score.
For the START, Vuln. = Vulnerability score and Str. = Strength score.

other hand was approximately one standard deviation below the means for male forensic patients and male offenders reported in the instrument manual (Hare, 2003).

2.3.1 Base Rate of Inpatient Violence

The overall sample initially included 99 unique individuals, but 7 files contained little or no usable information regarding outcomes, leaving 92 participants for predictive validity analyses. As noted above, sufficient data for coding each instrument or outcome variable were available for varying subsets of this sample, and thus the *n* values differ among analyses. For inpatient violence, outcome data were obtainable for 92 individuals and 28 individuals (30%) were identified as having engaged in inpatient violence; conversely, 70% of the sample did not engage in inpatient violence. On average, the patients were followed up for 1 year and 7 months between admission and discharge in hospital, although it is noted that follow-up times were quite variable (mean =19.37 months, SD = 26.97 months). A total of 11 study participants were never discharged from the hospital during the study period, while an additional 1 participant was deceased.

2.3.2 Interrater Reliability

Interrater reliability (Table 2.3) was assessed using intraclass correlation coefficients (ICCs) computed from 10 randomly selected cases. Interrater reliability was not assessed for the VRAG-R, because the VRAG-R was not scored independently by the study's research assistant. All ICCs were significant at $p < .05$. The START Strength scores at pretreatment produced the lowest ICC (.63), which was considerably lower than that for posttreatment scores (.89). It was noted that the observed discrepancy for one case appeared to be an outlier (Strength scores of 0 and 8, respectively), and so a new ICC was computed for START Strength pretreatment scores with the outlier removed, producing an alternative ICC of .84 ($p < .001$). Thus, it is possible that the ICC presented in Table 2.3 underestimates the interrater reliability of START Strength scores in the study.

2.3.3 Convergent and Discriminant Validity

Convergent and discriminant validity of risk measure scores were assessed in a number of ways. The instruments were correlated with one another at pretreatment (Table 2.4) and at posttreatment (Table 2.5); all instrument total scores were highly correlated ($p < .001$) at both time points. The lowest correlations were observed between the START Strength scores and the various risk measures, consistent with the separate, but related, constructs of risk and protective

Table 2.3

Interrater Reliability of Study Instruments: Single Measure Intraclass Correlation Coefficients (n = 10)

Instrument	Intraclass Correlation
HCR-20 Presence Total (PreTx)	.86***
HCR-20 Presence Total (PostTx)	.73**
HCR-20 Relevance Total (PreTx)	.87***
HCR-20 Relevance Total (PostTx)	.78**
START Vulnerability (PreTx)	.87***
START Vulnerability (PostTx)	.84***
START Strength (PreTx)	.63*
START Strength (PostTx)	.89***
VRS Total (PreTx)	.95***
VRS Total (PostTx)	.94***
PCL-R	.95***

Note. For all measures, PreTx = Pretreatment and PostTx = Posttreatment.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 2.4

Intercorrelations among Pretreatment Total Scores

	HCR-20 Rel.	HCR-20 SPJ C.P.	START V	START S	VRS	VRAG-R[a]	PCL-R[a]
HCR-20 Pres.	.86*** (88)	.69*** (89)	.77*** (88)	-.55*** (87)	.77*** (77)	.69*** (75)	.67*** (76)
HCR-20 Rel.		.78*** (88)	.61*** (87)	-.42*** (87)	.81*** (77)	.66*** (74)	.65*** (76)
HCR-20 SPJ C.P.			.45*** (89)	-.42*** (81)	.76*** (77)	.56*** (76)	.59*** (77)
START V				-.68*** (90)	.51*** (76)	.42*** (76)	.48*** (77)
START S					-.40*** (76)	-.39*** (75)	-.36*** (77)
VRS						.85*** (73)	.85*** (74)
VRAG-R							.83*** (75)

Note. *n* values vary by analysis and are provided in parentheses within cells containing under *r* values.

[a]VRAG-R and PCL-R were only scored once, at pretreatment.

For the HCR-20, Pres. = Presence score, Rel. = Relevance score, SPJ C.P. = Structured Professional Judgement Summary Statement for Case Prioritization. For the START, V = Vulnerability score and S = Strength score.

***$p < .001$.

Table 2.5

Intercorrelations among Posttreatment Total Scores

	HCR-20 Rel.	HCR-20 SPJ C.P.	START V	START S	VRS	VRAG-R[a]	PCL-R[a]
HCR-20 Pres.	.82*** (82)	.80*** (81)	.80*** (81)	-.68*** (81)	.84*** (73)	.79*** (73)	.75*** (74)
HCR-20 Rel.		.77*** (80)	.56*** (80)	-.48*** (80)	.75*** (73)	.63*** (72)	.56*** (74)
HCR-20 SPJ C.P.			.64*** (80)	-.55*** (80)	.77*** (72)	.66*** (72)	.66*** (73)
START V				-.74*** (81)	.67*** (72)	.59*** (72)	.59*** (73)
START S					-.52*** (72)	-.41*** (72)	-.38*** (73)
VRS						.86*** (70)	.85*** (71)
VRAG-R							.83*** (75)

Note. *n* values vary by analysis and are provided in parentheses within cells containing under *r* values.

[a]VRAG-R and PCL-R were only scored once, at pretreatment.

For the HCR-20, Pres. = Presence score, Rel. = Relevance score, SPJ C.P. = Structured Professional Judgement Summary Statement for Case Prioritization. For the START, V = Vulnerability score and S = Strength score.

***$p < .001$.

factors. Additionally, various static and dynamic subcomponents of the risk measures were correlated with one another, to assess convergent and discriminant validity, at pretreatment (Table 2.6) and posttreatment (Table 2.7). All the measures were significantly correlated with one another, with the exception of VRS Static scores and HCR-20^{V3} Clinical scores at pretreatment (the same correlations were low, but significant at posttreatment). Notably, these findings demonstrate at least some evidence of discriminant validity among the static components of the VRS and the dynamic/clinical components of the HCR-20^{V3} – a pattern that was less evident among the respective static and dynamic components of the individual tools. Additionally, the VRS Dynamic scores were more highly correlated with HCR-20^{V3} Historical factors than Clinical factors, indicating that they are not capturing equivalent constructs. Indeed, the pattern of intercorrelations observed among the various measures suggested that the historical/static and clinical/dynamic components of particular tools are not distinct from each other, and also that the clinical/dynamic components of different tools are not necessarily capturing identical constructs.

2.3.4 Research Objective 1: Detecting Dynamic Risk: The VRS, START, and HCR-20^{V3}

To evaluate the capacity of dynamic instrument scores to detect treatment related changes in risk, *t*-tests were conducted (Table 2.8). Cohen's *d* effect sizes were also computed to provide an indication of the magnitude of changes in risk, with effects of .20 indicating a small effect, .50 indicating a medium effect, and .80 indicating a large effect (Cohen, 1992). The results supported the hypotheses related to the HCR-20^{V3} Clinical scale, with the Presence (t (81) = 6.55, p <.001) and the Relevance scores (t (81) = 3.91, p < .001) exhibiting medium sized effects. START Vulnerability scores also produced significant changes (t (80) = 7.63, p < .001), with a medium effect. It was hypothesized that the VRS would not detect changes; however, VRS Dynamic scores produced statistically significant changes (t (72) = 6.55, p < .001), although the magnitude of these changes were less than Cohen's criteria for a small effect.

Beyond the study hypotheses, further exploratory *t*-tests were conducted to explore the ability of the other dynamic instruments' total scores and respective subscales to detect changes in risk. With the exception of the HCR-20^{V3} Historical scale and the VRS Static total, each of these analyses produced small but statistically significant effects.

Table 2.6

Intercorrelations among Dynamic and Static Risk Measure Components at Pretreatment

	HCR-20 C Pres.	HCR-20 R Pres.	HCR-20 H Rel.	HCR-20 C Rel.	HCR-20 R Rel.	START V	VRS Stat.	VRS Dyn.
HCR-20 H Pres.	.32*** (90)	.61*** (89)	.89*** (89)	.39*** (90)	.60*** (89)	.53*** (89)	.76*** (77)	.81*** (76)
HCR-20 C Pres.		.55*** (89)	.28** (89)	.85*** (90)	.49*** (89)	.72*** (89)	.02 (77)	.37*** (76)
HCR-20 R Pres.			.54*** (88)	.54*** (89)	.83*** (89)	.65*** (88)	.47*** (77)	.67*** (76)
HCR-20 H Rel.				.48*** (89)	.67*** (88)	.46*** (88)	.71*** (77)	.86*** (76)
HCR-20 C Rel.					.60*** (89)	.63*** (89)	.14 (77)	.49*** (76)
HCR-20 R Rel.						.56*** (88)	.40*** (77)	.65*** (76)
START V							.33*** (76)	.55*** (76)
VRS Stat.								.69*** (75)

Note. n values vary by analysis and are provided in parentheses within cells containing under *r* values.

For the HCR-20, H = Historical scale, C = Clinical scale, R = Risk Management scale, Pres. = Presence scores, and Rel. = Relevance Scores. For the START, V = Vulnerability score. For the VRS, Stat. = Static subtotal and Dyn. = Dynamic subtotal.

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

Table 2.7

Intercorrelations among Dynamic and Static Risk Measure Components at Posttreatment

	HCR-20 C Pres.	HCR-20 R Pres.	HCR-20 H Rel.	HCR-20 C Rel.	HCR-20 R Rel.	START V	VRS Stat.	VRS Dyn.
HCR-20 H Pres.	.37*** (82)	.70*** (82)	.87*** (81)	.45*** (82)	.54*** (82)	.54*** (81)	.75*** (74)	.74*** (73)
HCR-20 C Pres.		.55*** (82)	.36*** (81)	.61*** (82)	.51*** (82)	.65*** (81)	.24* (74)	.45*** (73)
HCR-20 R Pres.			.59*** (81)	.52*** (82)	.76*** (82)	.63*** (81)	.52*** (74)	.64*** (73)
HCR-20 H Rel.				.56*** (81)	.61*** (81)	.51*** (80)	.72*** (74)	.84*** (73)
HCR-20 C Rel.					.68*** (82)	.42*** (81)	.23* (74)	.54*** (73)
HCR-20 R Rel.						.55*** (81)	.43*** (74)	.66*** (73)
START V							.46*** (73)	.67*** (72)
VRS Stat.								.69*** (73)

Note. n values vary by analysis and are provided in parentheses within cells containing under *r* values.

For the HCR-20, H = Historical scale, C = Clinical scale, R = Risk Management scale, Pres. = Presence scores, and Rel. = Relevance Scores. For the START, V = Vulnerability score. For the VRS, Stat. = Static subtotal and Dyn. = Dynamic subtotal.

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

Table 2.8

Measurement of Dynamic Changes in Risk from Pretreatment to Posttreatment for the HCR-20^{V3}, START, and VRS

Instrument	*n*	Pretreatment *M* (SD)	Posttreatment *M* (SD)	Cohen's *d*	*t*
HCR-20					
Presence					
H	82	12.58 (4.08)	12.48 (4.25)	0.02	.90
C	82	5.12 (2.94)	3.02 (2.32)	0.79	7.34***
R	82	6.99 (1.93)	6.23 (2.25)	0.36	4.29***
Total	82	24.69 (7.27)	21.99 (7.07)	0.38	7.56***
Relevance					
H	81	10.97 (4.23)	10.99 (4.27)	-0.01	-.51
C	82	6.09 (2.84)	5.25 (2.69)	0.30	3.91***
R	82	7.24 (2.08)	6.68 (2.37)	0.25	3.34***
Total	81	24.01 (8.03)	22.54 (8.36)	0.18	2.90**
START					
Vulnerability	81	18.97 (9.05)	12.70 (7.56)	0.75	7.63***
Strength	81	5.12 (7.16)	8.68 (8.37)	-0.46	-5.15***
VRS					
Dynamic	73	30.99 (11.30)	30.15 (11.52)	0.07	6.55***
Static	73	7.18 (4.47)	7.04 (4.39)	0.03	3.18
Total	73	37.99 (14.82)	37.02 (14.81)	0.06	6.75***

Note. *$p < .05$, two-tailed. **$p < .01$, two-tailed. ***$p < .001$, two-tailed

2.3.5 Research Objective 2: Prediction of Inpatient Violence on Forensic Psychiatric Inpatient Units

Receiver Operating Characteristic (ROC) analyses were completed to evaluate the predictive validity of pretreatment scores produced by each of the study instruments for inpatient violence (Table 2.9). Magnitudes of the AUCs are described in this text using Rice and Harris' (2005) guidelines, with .56 representing a small effect, .64 representing a medium effect, and .71 representing a large effect. While it was hypothesized that scores from each of the instruments would be associated with inpatient violence, only total scores from the dynamic measures (HCR-20^{V3}, START, and VRS) produced significant AUCs; total scores on the VRAG-R and PCL-R did not produce significant AUCs, although the PCL-R total score approached significance (p = .08). The START Strength total and the VRS total score produced medium effects, while the HCR-20^{V3} Presence total, HCR-20^{V3} Relevance total, and START Vulnerability total produced large effects.

Further ROC analyses were conducted to explore the predictive validity of the various instruments' subscales and components for inpatient violence. The HCR-20^{V3} Historical scale Presence and Relevance scores both produced medium sized and significant effects, while the Clinical and Risk Management scales' Presence and Relevance scores all produced large and significant AUCs. Additionally, the HCR-20^{V3}'s SPJ summary risk statements were also subjected to ROC analyses, consistent with previous research; the Case Prioritization and the Imminent Violence statements each produced significant AUCs, of medium and large magnitude, respectively. The VRS Dynamic total produced a medium sized and significant AUC, but the Static total AUC did not reach significance. Of the analyses with the PCL-R facets and factors, only Facet 4 (Antisocial) and Factor 2 (Social Deviance) produced significant AUCs (both small in magnitude).

2.3.6 Research Objective 3: Contribution of Dynamic Risk Measures to the Prediction of Inpatient Violence

A series of logistic regression analyses were conducted to assess the relative contributions of the different instruments and their components to the prediction of inpatient violence. With regard to hypothesis 3.1, pretreatment dynamic scores from the HCR-20^{V3} and VRS were entered into hierarchical logistic regression analyses, after their respective static components, to assess their incremental validity for the prediction of inpatient violence. The

Table 2.9

Receiver Operating Characteristic Analyses for Inpatient Violence: Area Under the Curve (AUC) Statistics

Instrument	*n*	AUC	95% CI
HCR-20 PreTx[a]			
Presence			
H	89	.64*	(.52, .77)
C	90	.76***	(.65, .88)
R	89	.76***	(.66, .86)
Total	88	.76***	(.66, .86)
Relevance			
H	89	.64*	(.53, .76)
C	90	.72***	(.61, .84)
R	89	.72***	(.61, .83)
Total	88	.70***	(.58, .81)
SPJ			
CP	90	.68**	(.56, .79)
SH	90	.44	(.32, .56)
IV	90	.75***	(.64, .86)
START PreTx			
V	91	.76***	(.65, .87)
S	90	.69**	(.57, .80)
VRS PreTx			
Static	77	.58	(.43, .73)
Dynamic	76	.69**	(.56, .82)
Total	77	.68**	(.55, .81)
VRAG-R			
Total	76	.60	(.44, .76)
PCL-R			
Facet 1	77	.52	(.38, .67)
Facet 2	77	.62	(.48, .76)
Facet 3	77	.63	(.50, .76)
Facet 4	77	.66*	(.52, .80)
Factor 1	77	.60	(.45, .74)
Factor 2	77	.65*	(.52, .79)
Total	77	.63	(.49, .77)

Note. [a]PreTx = Pretreatment scores.
For the HCR-20: H = Historical Scale, C = Clinical scale, R = Risk Management Scale, SPJ = Structured Professional Judgment summary statements, CP = Case Prioritization, SH = Serious Harm, and IV = Imminent Violence.
For the START: V = Vulnerability total score and S = Strength total score.
*$p < .05$. **$p < .01$. ***$p < .001$.

results are displayed in detail in Table 2.10. Both of the HCR-20 Presence and Relevance Clinical scores incrementally predicted inpatient violence beyond their respective Historical scores (Presence e^B = 1.42 and Relevance e^B = 1.34). The VRS produced a similar pattern with the Dynamic score incrementally predicting inpatient violence, beyond the Static score (e^B = 1.09). These e^B values should be interpreted with the awareness that a one point change in score carries different meanings for each tool. To explain further, the HCR-20^{V3} Clinical scale has a range of only 10 points between the minimum and maximum scores, while the VRS Dynamic score has a range of 60 points, and so changes of one point on the respective tools should not be interpreted communicating equivalent changes in risk.

Next, to answer research question 3.2, dynamic scores from the HCR-20^{V3}, START, and VRS were entered into logistic regression analyses with the VRAG-R, to assess their unique and independent contributions to inpatient violence risk prediction. A detailed summary of these analyses is displayed in Table 2.11. Overall models combing the VRAG-R with each of the HCR-20^{V3} Clinical scale Presence and Relevance scores and START Vulnerability and Strength scores reached significance; the overall model combining the VRS Dynamic score with the VRAG-R did not significantly predict inpatient violence (χ^2 (2) = 5.34, p = .07). With the exception of the VRS Dynamic scale, each of the measures incrementally predicted inpatient violence beyond the VRAG-R (HCR-20 Clinical: Presence e^B = 1.40 and Relevance e^B = 1.33; START: Vulnerability e^B = 1.13 and Strength e^B = .81).

2.3.7 Research Objective 4: The Relationship between Dynamic Changes in Risk and Inpatient Violence

A series of hierarchical logistic regression analyses were conducted using change scores, to examine whether measured changes in risk would demonstrate incremental relationships with decreased inpatient violence, controlling for pretreatment scores or overall risk (Table 2.12). With regard to the HCR-20^{V3}, a number of models were analyzed, comparing Clinical scores, Risk Management scores, and summed Clinical and Risk Management scores (consistent with Wilson et al., 2013) with pretreatment scores on the respective scales and with total scores. Of the models assessing Presence scores, only Clinical scale change scores demonstrated an incremental relationship with decreased inpatient violence, beyond pretreatment clinical scale scores; however, it is noted that combined Clinical and Risk Management Presence scores

Table 2.10

Logistic Regression Analyses: Incremental Contributions of Dynamic Variables over Static Variables for the Prediction of Inpatient Violence with HCR-20 and VRS Pretreatment Scores

Predictor	$\Delta\chi^2$	*B*	*SE*	Wald	e^B	95% CI
HCR-20 Presence (*n* = 90)						
Step 1						
Historical		.14	.06	4.64*	1.15	(1.01, 1.30)
Constant		-2.58	.89	8.49**	.08	
Model χ^2 (1) = 5.08*						
Step 2	14.44***					
Historical		.09	.07	1.56	1.09	(.95, 1.26)
Clinical		.35	.10	12.11***	1.42	(1.16, 1.72)
Constant		-3.99	1.13	12.45***	.02	
Model χ^2 (2) = 19.52***						
HCR-20 Relevance (*n* = 89)						
Step 1						
Historical		.13	.06	5.11*	1.14	(1.02, 1.28)
Constant		-2.29	.73	9.78**	.10	
Model χ^2 (1) = 5.60*						
Step 2	8.44**					
Historical		.06	.07	.68	1.06	(.93, 1.21)
Clinical		.29	.11	7.47**	1.34	(1.09, 1.65)
Constant		-3.35	.91	13.50***	.035	
Model χ^2 (2) = 14.03***						
VRS (*n* = 75)						
Step 1						
Static		.07	1.50	1.50	1.07	(.96, 1.20)
Constant		-1.34	6.96	6.96**	.26	
Model χ^2 (1) = 1.51						
Step 2	6.14*					
Static		-.07	.08	.70	.93	(.79, 1.10)
Dynamic		.08	.04	5.51*	1.09	(1.01, 1.16)
Constant		-2.93	.89	10.82**	.05	
Model χ^2 (2) = 7.65*						

Note. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 2.11

Logistic Regression Analyses of the Relative Contributions of the VRAG-R and the Dynamic Components of the HCR-20, START, and VRS to the Prediction of Inpatient Violence

Model	*B*	*SE*	Wald	e^B	95% CI
1 *HCR-20 Presence* (*n* = 76)					
VRAG-R	.02	.01	1.39	1.02	(.99, 1.05)
Clinical	.34	.11	9.87**	1.40	(1.14, 1.73)
Constant	-2.80	.70	15.91***	.06	
Model χ^2 (2) = 14.18**					
2 *HCR-20 Relevance* (*n = 7*6)					
VRAG-R	.02	.01	1.21	1.02	(.99, 1.04)
Clinical	.28	.11	6.72**	1.33	(1.07, 1.64)
Constant	-2.80	.81	12.00***	.06	
Model χ^2 (2) = 10.10**					
3 *START* (*n* = 76)					
VRAG-R	.00	.02	.01	1.00	(.97, 1.03)
Vulnerability	.13	.04	10.56***	1.13	(1.05, 1.22)
Constant	-3.52	.89	15.52***	.03	
Model χ^2 (2) = 15.17***					
4 *START* (*n* = 75)					
VRAG-R	.01	.01	.15	1.01	(.98, 1.03)
Strength	-.21	4.89	4.90*	.81	(.67, .98)
Constant	-.16	.18	.18	.85	
Model χ^2 (2) = 11.00**					
5 *VRS* (*n* = 73)					
VRAG-R	-.00	.02	.02	1.00	(.96, 1.04)
Dynamic	.06	.04	2.18	1.06	(.98, 1.15)
Constant	-2.75	1.31	4.4*	.06	
Model χ^2 (2) = 5.34					

Note. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 2.12

Logistic Regression Analyses: Evaluation of the Incremental Validity of Dynamic Risk Change Scores and Inpatient Violence

Model		*n*	*B*	*SE*	Wald	e^B	95% CI
1	*HCR-20* (Presence)	82					
	Block 1						
	C + R TxΔ[a]		.04	.07	.43	1.05	(.92, 1.19)
	Block 2						
	C + R TxΔ		-.03	.08	.11	.98	(.84, 1.13)
	PreTx[b] Total		.19	.05	12.20***	1.21	(1.09, 1.34)
2	*HCR-20* (Relevance)	81					
	Block 1						
	C + R TxΔ		-.17	.11	2.52	.85	(.69, 1.04)
	Block 2						
	C + R TxΔ		-.20	.11	3.18	.82	(.66, 1.02)
	PreTx Total		.11	.04	7.91**	1.12	(1.03, 1.21)
3	*HCR-20* (Presence)	82					
	Block 1						
	C + R TxΔ		.04	.07	.43	1.05	(.92, 1.19)
	Block 2						
	C + R TxΔ		-.19	.10	3.76†	.83	(.69, 1.00)
	PreTx C + R		.46	.11	16.26***	1.58	(1.27, 1.98)
4	*HCR-20* (Relevance)	82					
	Block 1						
	C + R TxΔ		-.16	.11	2.42	.85	(.69, 1.04)
	Block 2						
	C + R TxΔ		-.33	.14	5.92*	.72	(.55, .94)
	PreTx C + R		.33	.09	12.98***	1.39	(1.16, 1.66)
5	*HCR-20* (Presence)	82					
	Block 1						
	C TxΔ		.05	.09	.27	1.05	(.87, 1.26)
	Block 2						
	C TxΔ		-.11	.11	.97	.90	(.73, 1.11)
	PreTx Total		.20	.06	12.69***	1.22	(1.10, 1.37)
6	*HCR-20* (Relevance)	81					
	Block 1						
	C TxΔ		-.08	.13	.39	.93	(.72, 1.18)
	Block 2						

	C TxΔ		-.13	.14	.95	.88	(.67, 1.14)
	PreTx Total		.11	.04	7.77**	1.12	(1.03, 1.20)
7	*HCR-20* (Presence)	82					
	Block 1						
	R TxΔ		.08	.14	.35	1.09	(.82, 1.44)
	Block 2						
	R TxΔ		.13	.16	.67	1.14	(.84, 1.54)
	PreTx Total		.19	.06	12.13***	1.21	(1.09, 1.35)
8	*HCR-20* (Relevance)	81					
	Block 1						
	R TxΔ		-.66	.33	3.98*	.52	(.27, .99)
	Block 2						
	R TxΔ		-.61	.33	3.33	.54	(.28, 1.05)
	PreTx Total		.10	.04	6.73	1.11	(1.03, 1.20)
9	*HCR-20* (Presence)	82					
	Block 1						
	C TxΔ		.05	.09	.27	1.05	(.87, 1.26)
	Block 2						
	C TxΔ		-.37	.14	6.86**	.69	(.53, .91)
	PreTx C		.61	.15	16.40***	1.83	(1.37, 2.45)
10	*HCR-20* (Relevance)	82					
	Block 1						
	C TxΔ		-.08	.13	.35	.93	(.73, 1.19)
	Block 2						
	C TxΔ		-.35	.16	4.57*	.71	(.52, .97)
	PreTx C		.42	.12	12.35***	1.52	(1.21, 1.93)
11	*HCR-20* (Presence)	82					
	Block 1						
	R TxΔ		.08	.14	.35	1.09	(.82, 1.44)
	Block 2						
	R TxΔ		-.05	.16	.12	.95	(.70, 1.29)
	PreTx R		.84	.24	11.87**	2.32	(1.44, 3.74)
12	*HCR-20* (Relevance)	82					
	Block 1						
	R TxΔ		-.65	.33	3.90*	.52	(.27, 1.00)
	Block 2						
	R TxΔ		-.90	.39	5.39*	.41*	(.19, .87)

	PreTx R		.66	.21	10.17**	1.94**	(1.29, 2.92)
13	*START*	81					
	Block 1						
	Vulnerability TxΔ		.01	.03	.11	1.01	(.95, 1.08)
	Block 2						
	Vulnerability TxΔ		-.13	.05	6.97**	.88	(.80, .97)
	PreTx Vulnerability		.21	.05	15.44***	1.23	(1.11, 1.37)
14	*VRS*	73					
	Block 1						
	Dynamic TxΔ		-.81	.34	5.87*	.44	(.23, .86)
	Block 2						
	Dynamic TxΔ		-.73	.3	4.74*	.48	(.25, .93)
	PreTx Total		.04	.02	3.82*	1.04	(1.00, 1.08)

Note. [a]TxΔ = Treatment Change. [b]PreTx = Pretreatment score.
For the HCR-20: H = Historical scale, C = Clinical scale, R = Risk Management scale, C + R = Sum of Clinical and Risk Management scales.
[†]$p < .06$. *$p < .05$. **$p < .01$. ***$p < .001$.

approached significance (p = .052), controlling for pretreatment scores on those scales. HCR-20^{V3} Relevance change scores on the other hand, demonstrated incremental relationships with decreased inpatient violence in the following instances: combined Clinical and Risk Management change scores, controlling for pretreatment scores on these scales; Clinical scale change scores, controlling for pretreatment Clinical scores; and Risk Management change scores, controlling for pretreatment Risk Management scores. It is also noted that additional Relevance change scores approached significance in their associations with decreased inpatient violence, including combined Clinical and Risk Management Presence change scores, controlling for pretreatment total scores (p = .07), and Risk Management scores, controlling for pretreatment total scores (p = .07).

Additionally, START Vulnerability change scores demonstrated an incremental relationship with decreased inpatient violence, beyond pretreatment START scores (e^B = .88, p < .01), and VRS Dynamic change scores demonstrated an incremental relationship with decreased inpatient violence, beyond pretreatment total scores (e^B = .48, p < .05).

2.4 Discussion

2.4.1 Research Objective 1: Detecting Dynamic Risk: The VRS, START, and HCR-20^{V3}

With the goal of assessing the instruments' utility for detecting dynamic changes in risk, a series of *t*-tests were conducted. With the exception of the analysis with the VRS, which produced a small but statistically significant reduction in mean risk score (less than one point), the hypotheses regarding change measurement were generally supported. It is also noted that despite a significant *t*-statistic, the VRS change analysis produced a Cohen's *d* of .07 for the Dynamic item total, indicating that the observed differences likely have little practical utility for dynamic risk assessment for inpatient violence in this sample. The HCR-20^{V3} Clinical and Risk Management scales, as well as the START Strength and Vulnerability total scores, all captured small to medium sized change effects, based on Cohen's *d* values. The HCR-20^{V3} Clinical scale Presence score produced the largest change effect, with a Cohen's *d* of .79 (.80 is the cut-off for a "large" effect). Given that AHE is a psychiatric hospital, it is not surprising that the scales capturing more clinical variables (e.g., mental health symptomatology and instability) demonstrated greater change effects. It should also be noted that due to constraints imposed by the available data, these change analyses reflect treatment related changes that occurred over variable follow-up periods.

The observed difference between the change effects produced by the HCR-20^{V3} Clinical scale Presence and Relevance scores (Cohen's *d* values of .79 and .30, respectively) will likely be of interest to some readers. This discrepancy makes sense when considered along with the distinction between these two scores, which is a new feature of the third version of the HCR-20 scheme. Consider, for example, the meaning of risk scores on the Symptoms of Major Mental Disorder item, in a forensic psychiatric hospital like AHE. As explained previously, individuals who are found NCR on account of a mental disorder may be detained for an indeterminate period, with factors such as public safety and mental state influencing release decisions. As a result, it is unlikely that patients will be discharged until they demonstrate a reduction in the presence of acute symptoms of their mental disorders, which should be reflected in low posttreatment Presence scores on the corresponding HCR-20^{V3} item. Coding of Relevance scores on the other hand, should communicate important information about the historical and/or future implications of the variable, including its relative importance to ongoing risk management decisions, which may not be captured in Presence scores. In the case of a patient or offender who has demonstrated a longstanding pattern of violent behaviour precipitated by deteriorations in mental state, it makes sense that reductions in Presence scores would not necessarily be matched by corresponding reductions in Relevance scores. Thus, while HCR-20^{V3} Presence and Relevance scores correspond to the same risk factors, the current findings support the contention that they capture distinct empirical and theoretical dimensions of these factors.

2.4.2 Research Objective 2: Prediction of Inpatient Violence on Forensic Psychiatric Inpatient Units

The primary aim of this program of research was to determine the predictive validity of existing instruments for inpatient violence. As mentioned above, analyses were influenced by the raw data available in institutional files. In particular, it was not possible to reliably identify exact dates for all events described within the files. This precluded the recoding of instruments at fixed intervals, and so predictive validity analyses for inpatient violence were limited to pretreatment scores and a single comprehensive follow-up period (i.e., any aggression during the index admission). Thus, the current ROC analyses may be interpreted as assessing the pretreatment scores' predictive validity for any inpatient aggression during complete admissions of varying lengths.

Generally speaking, the results of the various ROC analyses appear to lend some empirical support to the argument that structured risk instruments designed to capture dynamic/clinical variables are particularly well suited to the assessment of risk for inpatient violence (Daffern, 2007). The measures designed to capture these types of variables, such as the START Vulnerability score and the HCR-20^{V3} Clinical scale, produced the largest predictive effects, while measures that were not designed to capture these variables, such as the HCR-20^{V3} Historical scale and the VRAG-R, produced smaller effects.

This pattern of more dynamic/clinical risk measures demonstrating stronger predictive accuracy than more static/stable measures was also observed among the components of the VRS, with the Dynamic score performing better than the Static score in this study. These findings are quite consistent with the limited previous research using the VRS to assess risk for inpatient violence, which also demonstrated greater predictive accuracy for the Dynamic score (Dolan & Fullam, 2007; Dolan et al., 2008). In addition, the two studies conducted by Dolan and Fullam (2007) and Dolan and colleagues (2008) produced comparable predictive effects to the current study overall, with Dynamic and Total score effects equivalent to AUCs ranging from .69 to .71 (Rice & Harris, 2005), which are similar to the current AUCs of .69 and .68. Overall, these ROC analyses supported the predictive validity of the VRS for forensic psychiatric inpatient aggression.

The HCR-20 Version 3 contains a number of elements which each command individual attention. While this is a new version of an established scheme and thus has yet to receive extensive study, the previous iterations have shown promise for the prediction of inpatient violence (Chu et al., 2013; Wilson et al., 2013). In their meta-analysis of inpatient violence prediction research, Hogan and Ennis (2010) reported overall predictive effects equivalent to AUCs of .69 for Total scores and .70 for the Clinical scale. In the current study, various components of the updated version of the HCR-20 appeared to either match (e.g., Relevance Total and Clinical scores, Case Prioritization summary statement: AUCs of .70, .72, and .68) or exceed these effects (e.g., Presence Total and Clinical scores, Imminent Violence summary statement: AUCs of .76, .76, and .75). It is also worth noting that summed scores, which are commonly used in the HCR-20 research literature (Douglas et al., 2014) and the SPJ summary statements that the authors recommend for clinical practice, each demonstrated high levels of

predictive accuracy for inpatient violence. These results indicate that much like its predecessors, the HCR-20^{V3} appears to be well suited to the assessment of risk for inpatient violence.

The START was specifically designed to measure dynamic variables which may impact risk (Webster et al., 2009) and thus is also theoretically well-suited to the assessment of risk for inpatient violence. While acknowledging that the existing research base was small, Chu and colleagues (2011) presented data from a number of early evaluations of the START Vulnerability scores across a variety of settings, with AUCs for the prediction of any aggression ranging from .52 to .82. Subsequent studies using forensic psychiatric samples, including Chu and colleagues (2011) own results, have produced AUC values ranging from .76 to .82 for Vulnerability scores, and from .71 to .84 for Strength scores (Desmarais et al., 2012; Wilson et al., 2013). In this study, both the Vulnerability and Strength total scores predicted inpatient violence with AUC values of .76 (large) and .69 (medium), respectively. These results represent further empirical support for the use of the START, and the Vulnerability score in particular, as a measure of risk for inpatient violence in forensic psychiatric settings.

As mentioned above, the VRAG-R total score did not significantly predict inpatient violence based on ROC analyses in this study. Notably, the small and insignificant AUC effect observed for the VRAG-R does not appear to be consistent with the limited research supporting the use of the VRAG for the prediction of inpatient violence. For example, previous studies conducted by Doyle, Dolan, and McGovern (2002) and Vitacco and colleagues (2012b) both reported predictive effect sizes equivalent to AUCs of .71. Given the limited number of studies using the VRAG and the VRAG-R for the prediction of inpatient violence, further research will be necessary to establish the tool's predictive validity for this outcome.

Like the VRAG-R, the PCL-R total score did not produce a significant predictive effect for inpatient violence in this study. However, it is also noted that the lack of statistical significance observed for the PCL-R total score may have been influenced by statistical power limitations, given that the observed AUC of .63 is consistent with the result reported in Hogan and Ennis' (2010) meta-analysis, which was equivalent to an AUC of approximately .62; the observed effect was also consistent with Yang, Wong, and Coid's (2010) reported overall effect of .64, which was computed from a more diverse selection of forensic samples. While the total score did not produce a significant effect in this study, the Antisocial Facet and Social Deviance Factor did produce medium sized and significant effects. It is also worth acknowledging that the

PCL-R was not designed to assess risk per se, and that its use may confer benefits that were not measured in this study, such as the identification of responsivity needs that can guide treatment services (Andrews & Bonta, 2010; Olver & Wong, 2011). Nonetheless, given the current and previous findings, it is likely advisable to use a purpose-built instrument, rather than the PCL-R, to assess risk for inpatient violence. Assessors interested in using a measure of psychopathy may also consider the use of the PCL:SV as an alternative to the PCL-R, given that it has produced larger effects in the existing literature (Hogan & Ennis, 2010) and the fact that administration of the PCL:SV is less time-intensive.

2.4.3 Research Objective 3: Contribution of Dynamic Risk Measures to the Prediction of Inpatient Violence

The trend of clinical/dynamic measures producing the strongest predictive effects for inpatient violence continued with these logistic regression analyses. Furthermore, each of the HCR-20^{V3} Presence and Relevance Clinical scores incrementally predicted inpatient violence beyond their respective Historical scores; controlling for Historical scores, one point increases in Clinical Presence and Relevance scores were associated with increases of 42% and 34% in the odds of inpatient aggression, respectively. The VRS produced a similar pattern, with a one point increase in Dynamic score associated with a 9% increase in the odds of inpatient aggression. Similar logistic regression analyses were then conducted to compare dynamic and static variables, using the VRAG-R as a common measure of static risk. With the exception of the VRS Dynamic scale, each of the measures incrementally predicted inpatient violence beyond the VRAG-R. Controlling for VRAG-R scores, one point increases in HCR-20 Clinical Presence and Relevance scores were associated with increases of 40% and 34% in the odds of inpatient aggression, while one point increases in START Vulnerability and Strength scores were associated with an increase of 13% and a decrease of 19% in the odds of inpatient aggression, respectively. Thus, clinical/dynamic factors made a significant contribution to the prediction of inpatient violence in this study.

This is not the first study to compare static and dynamic variables' incremental contributions to the prediction of inpatient violence, although previous comparisons have been conducted using different instruments. Based on null hypothesis significance testing alone, the results of these analyses appear to have produced conflicting results. Desmarais and colleagues (2012) found that START Vulnerability scores demonstrated incremental prediction beyond the

HCR-20 Version 2 Historical scale for the prediction of inpatient aggression. Similarly, Wilson and colleagues (2013) summed the HCR-20 Version 2 Clinical and Risk Management scales, and found that this combination incrementally predicted inpatient aggression beyond the Historical scale. They further assessed the incremental validity of the START beyond the HCR-20 Historical scale, and found that the Vulnerability scores, but not Strength scores, incrementally predicted the outcome variable. In contrast, Vitacco and colleagues (2012b) found that the HCR-20 Version 2 Clinical and Risk Management scores did not demonstrate incremental prediction beyond a collection of static measures, including the Historical scale, VRAG, and PCL-R. Looking beyond statistical significance however, the Clinical scale effect in Vitacco and colleagues' study was in the expected direction (e^B = 1.36). Furthermore, conclusions about these negative findings should be made with the caveat that the investigators used several different measures simultaneously as static controls, which were highly correlated with the dynamic measures – effectively decreasing the likelihood of detecting unique and independent effects. Overall then, the fact that dynamic variables did demonstrate statistically significant incremental prediction for inpatient violence in the current study is considered to be consistent with previous research as a whole.

In this study, the overall models combining both types of variables were found to be significant predictors of inpatient aggression. This being said, the instruments incorporating dynamic factors performed best, and dynamic factors made unique and independent contributions to the prediction of inpatient aggression, beyond static factors. Thus, while it is likely advisable to consider both static and dynamic variables in the assessment of risk for inpatient aggression, dynamic factors appear to warrant special attention.

2.4.4 Research Objective 4: The Relationship between Dynamic Changes in Risk and Inpatient Violence

There is often considerable empirical and conceptual overlap among risk factors that are identified as static and dynamic, respectively. As a result, an important limitation of many evaluations of the relative contributions of static and dynamic risk factors, which also applies to the current research, is that ostensibly dynamic variables tend to be highly correlated with static variables, which makes the detection of unique and independent prediction difficult. Indeed, significant correlations were observed among the scores of the current study measures. Thus, as was also discussed earlier in this document, evaluating actual change scores is a critical aspect of

the validation of dynamic risk factors (Douglas & Skeem, 2005; Kraemer et al., 1997; Olver et al., 2007). As mentioned above, the data available in institutional files precluded an assessment of the predictive validity of change scores for subsequent violence; instead, the concurrent validity of change scores and inpatient violence was assessed.

A number of HCR-20^{V3} change scores demonstrated incremental concurrent validity with inpatient violence, controlling for pretreatment scores on the respective scales. Risk Management Relevance change scores demonstrated the strongest relationship with inpatient violence, with a one point increase (i.e., indicating reduced risk) associated with a 59% reduction in the odds of inpatient aggression, controlling for pretreatment score. One point increases in change scores on each of the Clinical scale scores (Presence and Relevance), as well as the combined Clinical and Risk Management Relevance scores, were associated with reductions in the odds of inpatient aggression of between 28% and 31%. Of particular interest, overall, Relevance scores demonstrated stronger incremental relationships than Presence scores – alternative measures of the same risk variables.

START Vulnerability change scores demonstrated incremental concurrent validity with inpatient violence, controlling for pretreatment scores (e^B = .88); put another way, a one point increase in change score was associated with a 12% reduction in the odds of inpatient aggression. Despite capturing only limited treatment related changes overall (Cohen's d = .07), the VRS' dynamic change scores also demonstrated incremental concurrent validity with inpatient violence, beyond pretreatment total scores. The e^B value of .48 indicated that a one point increase in change score was associated with 52% reduction in the odds of inpatient aggression. Notably, while the earlier results suggested that few patients demonstrated treatment-related risk reductions as measured by the VRS (Cohen's d = .07), these results indicated that the individuals who did demonstrate risk changes also demonstrated a significantly lower probability of inpatient aggression.

Overall, change scores produced by each of the study's dynamic risk instruments demonstrated incremental concurrent validity with inpatient violence, beyond pretreatment scores. Specifically, positive risk change was associated with decreased violence over the course of the patients' stays, bearing in mind the uncertainty of the timing of aggressive incidents (i.e., whether they preceded or followed evaluations of change) and that change ratings were made during the latter part of the patient's stay during the study period. While predictive or causal

relationships cannot be definitively inferred from these concurrent validity findings, the data provided some evidence suggesting that the measurement of dynamic changes in violence risk may well be a clinically viable and empirically justifiable practice. Further research on the predictive validity and dynamic nature of these risk measures will be necessary to evaluate whether they can ultimately fulfill Kraemer and colleagues' (1997) criteria for causal dynamic risk factors – elsewhere described as the "holy grail … of offender change" (Serin et al., 2013, p. 32).

Additionally, the results of the HCR-20^{V3} analyses also provided some support for the idea that assessing the ongoing risk-relevance of clinical/dynamic factors (e.g., acute psychiatric symptoms) requires more than simply measuring their presence or absence. Indeed, while they were not compared directly, the superior performance of HCR-20^{V3} Relevance scores over Presence scores may be interpreted as reflecting a potentially important distinction, between reliable, clinically and practically relevant changes (Olver & Wong, 2011), and transitory fluctuations, or changes that are dependent upon external factors, such as restrictions, supervision, or other temporary risk management strategies. This point is perhaps best illustrated by an example. Consider an individual with a history of violent incidents precipitated by poor medication adherence and subsequent deteriorations in his mental state. At discharge from a forensic psychiatric institution, this individual is likely to receive a low Presence rating on the Symptoms of Major Mental Disorder item of the HCR-20^{V3}, even if medication adherence is externally motivated. Taken alone, this Presence score could wrongly suggest to the recipients of a risk assessment report that acute symptoms of psychosis are not a key risk management consideration for this individual, whereas a Relevance score should continue to emphasize the importance of monitoring and managing of his symptoms. The HCR-20^{V3}'s Relevance scores may provide an empirically valid means of building a case conceptualization/formulation (Hart, Sturmey, Logan & McMurran, 2011; Lewis & Doyle, 2009), that informs ongoing intervention and risk management decisions.

2.4.5 Strengths and Limitations

This study had some notable strengths and limitations that are worth acknowledging. By assessing the predictive validity of a number of new and existing instruments for forensic psychiatric inpatient violence, this study addressed a need in the literature. It is also worth noting that to the author's knowledge at the time of writing, there were no published studies of

the predictive validity of the HCR-20^{V3} or VRAG-R for forensic psychiatric inpatient violence.

By selecting consecutive admissions to a forensic psychiatric hospital as participants, the sample can be considered highly representative of the population served by AHE. Furthermore, given that the sample composition also appeared to be similar to the composition of the Canadian forensic psychiatric population more generally (Latimer and Lawrence, 2006), it is likely that the results will also be relevant to other Canadian forensic mental health programs. On the other hand, the presence of the small group of offenders referred for DO/LTO assessments is not representative of all forensic psychiatric populations across Canada, and may have served to increase the risk profile of the sample, relative to similar hospitals. Additionally, the findings represent a contribution to the literature on the relationship between dynamic risk assessments and inpatient violence.

There were also notable limitations associated with this study. For example, like all archival research, the results were certainly limited by the data contained in institutional files. This may be considered a limitation of not only the scoring of the risk instruments, which could not be completed for a number of patients, but also of the collection of the outcome data, given that institutional incidents are often underreported (Woods et al., 2008; Woods et al., 2015). Furthermore, given that the timing of incidents of inpatient violence could not be reliability identified, the current results do not provide specific information about the predictive validity of study instruments over specific time frames, which previous studies have provided (e.g., Dolan & Davies, 2006; Gray et al., 2003; Vitacco et al., 2012b). This lack of specific time information limits the extent to which the study findings can be used to infer predictive validity of actual change scores for inpatient violence. Thus, the paradigm employed in this study, which linked measurable changes in risk to concurrent inpatient violence, did not allow the author to address the question of whether the study measures capture truly causal dynamic risk factors (Andrews & Bonta, 2010; Andrews et al., 2006; Douglas & Skeem, 2005; Kraemer et al., 1997). Finally, it should also be noted that the START analyses were based exclusively on summed total scores, rather than the SPJ summary statements the authors recommend for clinical practice.

2.5 Conclusion

The findings of this study produced valuable data to inform its primary research questions. A number of new and more established instruments received support for their predictive validity with regard to inpatient violence among forensic psychiatric patients. The

results further indicated that risk factors that have been theoretically identified as clinical/dynamic variables were incrementally predictive of inpatient aggression, beyond static factors. Additionally, indirect empirical support for the dynamic nature of the risk measures was provided through concurrent validity analyses. Thus, forensic mental health professionals may use these findings to assist them in selecting risk instruments, based on the particular goals and intended applications of individual assessments (Douglas, 2014; Yang et al., 2010).

Chapter 3: STUDY 2

3.1 Introduction

In keeping with the overall goals of this research, the current prospective study was conducted to evaluate the predictive validity of a number of forensic instruments for inpatient violence. Additionally, the relative contributions of components of the instruments to the assessment of risk for inpatient violence were also evaluated. Hypotheses and research questions were consistent with the objectives of this program of research as a whole.

3.1.1 Hypotheses/Research Questions

3.1.1.1 Research Objective 1: Detecting Dynamic Risk: The VRS, START, and HCR-20^{V3}.

1.1. Change scores, produced by computing the differences between scores completed 28 days apart, were predicted to be significantly greater than zero, for the following instruments: HCR-20^{V3} Clinical Scale, HCR-20^{V3} total score, START total score.

1.2. It was expected that change scores produced by the VRS Total Score would not be significantly greater than zero, due to the variability in formal violence reduction programming provided to participants.

3.1.1.2 Research Objective 2: Prediction of Inpatient Violence on Forensic Psychiatric Inpatient Units.

2.1. It was anticipated that higher risk scores, computed from the HCR-20^{V3}, PCL-R, START, VRS, and VRAG-R, would be associated with a greater probability of a violent incident within the institution over the follow-up period(s).

2.2. Using Receiver Operating Characteristic (ROC) analyses, it was expected that HCR-20^{V3}, PCL-R, START, VRS, and VRAG-R scores would demonstrate predictive accuracy (i.e., AUC values significantly greater than .50) with respect to outcome measures of inpatient violence.

3.1.1.3 Research Objective 3: Contribution of Dynamic Risk Measures to the Prediction of Inpatient Violence.

3.1. It was expected that dynamic factors of the VRS and HCR-20^{V3} Clinical Scale would demonstrate incremental validity for the prediction of institutional violence, after controlling for the static factors of the VRS and HCR-20^{V3} Historical Scale, respectively.

3.2. Dynamic scores, produced by the VRS, START, and the HCR-20 Clinical scale, were entered into logistic regression analyses along with a measure of static risk, the VRAG-R, to

determine the unique and independent contributions of each to the prediction of the dichotomous outcome variable (any violent incident in the follow-up period).

3.2 Method

3.2.1 Participants

The participants in Study 2 included a sample of adult forensic psychiatric inpatients, admitted to a maximum security unit in the Acute Assessment and Treatment Program, Forensic Psychiatry Services, Alberta Hospital Edmonton (AHE). The overall sample included three partially overlapping cohorts, with individuals followed up for at least one of three consecutive 28 day periods. To ensure adequate data for instrument scoring, only patients who had resided on the unit for a minimum of two weeks prior to one of the three follow-up periods were included, resulting in a total of 19 unique individuals.

A summary of demographic and other relevant participant characteristics is presented in Table 3.1. The mean age at the start of the respective follow-up periods was approximately 38.76 (SD = 13.76 years). The sample was 84% male and 16% female. A total of 11 individuals were identified as white, while 3 were identified as being of Canadian Aboriginal descent, and the remaining individuals were identified as being of other diverse ethnic backgrounds. The majority of the sample had not completed high school (84%). Approximately three quarters of the sample had been found NCR, while the remaining individuals had been found unfit to stand trial, were prisoners receiving mental health care, or were undergoing an assessment. Schizophrenia and related psychotic disorders were the most common diagnoses (69%) observed in this sample. A current or previous diagnosis of a substance or alcohol related disorder was also common (78%). With regard to criminal and violent histories, 82% of the sample had received at least one criminal charge prior to their index, and 95% of the sample had committed a violent index offence. The majority of the sample also had a history of institutional aggression (90%), in either a correctional or hospital setting.

3.2.2 Program Description

See Study 1 for a detailed summary of the Acute Assessment and Treatment Program at AHE. One relevant change to the program was noted between the time periods associated with Study 1 and Study 2: at the time Study 2 was conducted, the program no longer regularly admitted individuals for assessments related to Dangerous Offender (DO) or Long-Term

Table 3.1

Summary of Relevant Demographic, Mental Health, and Historical Variables

	Mean (SD)	%
Demographics		
Age at First Assessment (years)	38.76 (13.13)	
Male		84
Female		16
White		58
Aboriginal Descent		16
Education		
Less than Grade 12		84
High School Diploma/GED		5
Any Post-Secondary		11
Admission Status[a]		
NCR		74
Certified Prisoner		11
Unfit		11
Other		5
Mental Health		
Schizophrenia/Psychotic Disorder		69
Antisocial Personality Disorder		37
Substance/Alcohol Diagnosis		78
Historical		
Any Previous Criminal Charge		82
Violent Index		95
Previous Institutional Aggression		90

Note.n = 19.

[a]Percentages do not equal 100% due to rounding error.

Offender (LTO) designations, resulting in changes to the composition of the patient populations between the two studies.

3.2.3 Materials

Materials were consistent with those used in Study 1.

3.2.4 Procedure

Ethical and administrative approvals were obtained along with those required for Study 1 as described previously.

This was a prospective study, completed in three waves. All instruments were scored by the author. Risk assessments were conducted based on information available in current institutional files, with clinical/dynamic ratings emphasizing data available in the two weeks prior to a common assessment date. In order for the results to reflect real world constraints, every study instrument and subscale with sufficient data for scoring was included in analyses; as a result, some individuals received scores on only clinical measures like the HCR-20 Clinical scale or the START Vulnerability scale. Outcome data were then obtained from institutional files and shift reports at the end of the 28 day follow-up period, using the same operational definition as Study 1, based on the SOAS-R. This procedure was repeated for a total of three separate waves.

It is also noted that SOAS-R outcome rating forms were completed by clinical staff at AHE. However, due to a low completion rate and significant discrepancies between the SOAS-R forms and other incident reporting methods, analyzing the predictive validity of the study instruments with SOAS-R forms did not prove feasible.

3.2.5 Data Analytic Plan

For the purposes of the study analyses, each day for each individual patient was treated as a separate unit of analysis, in order to maximize statistical power. This type of approach has been employed in previous research on inpatient and institutional violence (Almvik, Woods, & Rasmussen, 2000; Chu et al., 2013; Wilson, Desmarais, Nicholls, & Brink, 2010; Wilson et al., 2013; Woods et al., 2015)

3.2.5.1 Preliminary Analyses: Descriptive Statistics and Convergent Validity. In addition to the computation of basic descriptive statistics (e.g., mean risk scores), convergent validity was assessed by computing correlations among instrument scores. For the purposes of these preliminary analyses, each unique monthly rating was treated as a unit of analysis.

3.2.5.2 Research Objective 1: Detecting Dynamic Risk: The VRS, START, and HCR-20^{V3}. Despite the limited number of participants remaining on the unit for two subsequent assessment periods, *t*-tests were conducted as planned on these cases to evaluate whether each dynamic instrument detected measurable changes between subsequent assessments.

3.2.5.3 Research Objective 2: Prediction of Inpatient Violence on Forensic Psychiatric Inpatient Units. ROC analyses producing area-under-the-curve (AUC) statistics were conducted on the HCR-20^{V3}, PCL-R, START, VRS, and VRAG-R monthly scores, with respect to inpatient violence. Scores on the risk instruments were coded as continuous variables, while inpatient violence was coded as a dichotomous outcome variable (any versus no incidents of inpatient violence during each individual day of the corresponding follow-up period). Total scores, subcomponents, and HCR-20^{V3} SPJ summary statements (Desmarais et al., 2012; Gray et al., 2011; Wilson et al., 2013) were each analyzed in this manner. In addition to combined analyses using data from all three follow-up periods, separate analyses were also conducted for each wave of the study.

3.2.5.4 Research Objective 3: Contribution of Dynamic Risk Measures to the Prediction of Inpatient Violence. Using the daily dichotomous inpatient violence outcome variable, hierarchical logistic regression analyses were conducted to determine whether the dynamic scales demonstrated incremental validity over their corresponding historical/static scales. Logistic regression analyses were also used to assess the unique and independent contributions of the various dynamic scores compared with the VRAG-R.

3.3 Results

As explained in the description of the study procedure, every measure that could be scored, including subscales, was included in the analyses. Thus, the sample sizes vary among the various analyses. Table 3.2 contains a summary of the sample size, mean, and standard deviation of scores on each instrument across the entire study, while Table 3.3 contains similar data broken down by study waves.

Generally speaking, the distributions of scores on the risk instruments were higher in this study than the distributions observed in Study 1, despite the absence of offenders referred for DO/LTO assessments. The mean PCL-R score in this study was comparable to the means reported for male forensic patients and male offenders in the instrument manual (Hare, 2003). The mean START Vulnerability score was approximately one standard deviation higher than the

Table 3.2

Descriptive Statistics: Means and Standard Deviations of Instrument Scores

Instrument	*n*	Mean	SD
HCR-20 H Pres.	33	14.79	4.06
HCR-20 C Pres.	34	6.32	3.21
HCR-20 RM Pres.	34	8.12	1.27
HCR-20 Total Pres.	33	29.09	6.44
HCR-20 H Rel.	33	12.70	3.96
HCR-20 C Rel.	34	6.62	3.01
HCR-20 RM Rel.	34	8.21	1.20
HCR-20 Total Rel.	33	27.42	6.14
START Vuln.	33	25.40	7.98
START Str.	33	2.53	3.00
VRS Static	33	8.88	3.88
VRS Dynamic	33	40.74	10.35
VRS Total	33	49.57	12.78
VRAG-R	33	11.25	19.11
PCL-R	31	18.01	11.29

Note. PreTx = Pretreatment score. PostTx = Posttreatment score.
For the HCR-20, H = Historical scale, C = Clinical scale, R = Risk Management scale, Pres. = Presence score, Rel. = Relevance score.
For the START, Vuln. = Vulnerability score and Str. = Strength score.

Table 3.3

Descriptive Statistics: Means and Standard Deviations of Instrument Scores by Study Wave

	Wave 1			Wave 2			Wave 3	
Instrument	*n*	Mean	SD	*n*	Mean	SD	*n*	Mean	SD
HCR-20 Historical Scale Pres.	12	14.17	4.24	11	14.91	3.94	10	15.40	4.30
HCR-20 Clinical Scale Pres.	12	5.67	3.03	11	6.09	3.70	11	7.27	2.94
HCR-20 Risk Management Scale Pres.	12	8.25	1.29	11	8.09	1.38	11	8.00	1.26
HCR-20 Total Pres.	12	28.08	7.24	11	29.09	6.01	10	30.30	6.36
HCR-20 Historical Scale Rel.	12	12.25	4.07	11	12.73	3.90	10	13.20	4.26
HCR-20 Clinical Scale Rel.	12	6.75	2.67	11	5.91	3.48	11	7.18	2.99
HCR-20 Risk Management Rel.	12	8.58	1.16	11	8.09	1.30	11	7.91	1.14
HCR-20 Total Rel.	12	27.58	6.71	11	26.73	6.00	10	28.00	6.16
START Vulnerability	12	23.83	7.70	11	25.84	8.16	10	26.82	8.60
START Strength	12	3.10	3.93	11	2.02	2.42	10	2.42	2.42
VRS Static	12	8.08	4.42	11	8.82	3.08	10	9.90	4.14
VRS Dynamic	12	39.77	11.05	11	40.20	10.49	10	42.51	10.22
VRS Total	12	47.84	14.35	11	48.92	12.03	10	52.35	12.46
VRAG-R	12	6.54	21.57	11	11.84	17.12	10	16.25	18.63
PCL-R	10	17.50	12.70	11	17.45	10.64	10	19.14	11.65

Note. For the HCR-20, Pres. = Presence score and Rel. = Relevance score.

corresponding score observed in Study 1, as well as the scores reported in previous studies by Desmarais and colleagues (2012) and Wilson and colleagues (2013). The mean VRAG-R score fell between the 68th and 70th percentiles reported in the official scoring manual. The mean VRS total score was over half a standard deviation higher than the mean reported in Wong and Gordon's (2006) normative study of 918 offenders.

3.3.1 Base Rates of Inpatient Violence

While the sample included 19 unique individuals, there were a total of 34 separate assessments. Treating each day of the follow-up periods as a unique observation, there were 886 units of analysis over the three months of the study. In total 50 violent incidents were documented over the entire study period, corresponding to a base rate of 5.6% (only 5 of these incidents were documented with SOAS-R forms). Separate base rates were also calculated for each of the three waves of the study, and were observed to be 6.6%, 6.9%, and 3.6%, respectively. Thus, over 90% of all observations did not include an act of inpatient violence.

3.3.2 Convergent and Discriminant Validity

Convergent validity and discriminant validity of the various risk instruments were assessed by correlating them with one another. Of the total scores that were assessed (Table 3.4), all were significantly correlated with one another, except for the HCR-20^{V3} SPJ Imminent Violence summary rating, which was not significantly correlated with the VRAG-R or PCL-R. As expected, all risk measures were significantly negatively correlated with the START Strength score. The pattern of correlations among subscales was more complex (Table 3.5). While the majority of the subscales were significantly correlated, there were a few exceptions. HCR-20^{V3} Clinical Presence scores were not significantly correlated with HCR-20^{V3} Historical or Risk Management scores, or with VRS Static scores. HCR-20^{V3} Clinical Relevance scores were not significantly correlated with Historical Presence sores, Risk Management scores (Presence or Relevance), or VRS Static scores. Additionally, the HCR-20^{V3} Risk Management Relevance score was not significantly correlated with the VRS Static score.

3.3.3 Research Objective 1: Detecting Dynamic Risk: The VRS, START, and HCR-20^{V3}

A series of *t*-tests were conducted to determine whether the study instruments captured dynamic changes in risk when scored for the same individual during two consecutive assessment periods (Table 3.6). In total, there were 15 cases in which consecutive assessments were available. None of the analyses produced significant *t*-statistics. Cohen's *d* values were very

Table 3.4

Intercorrelations among Study Instruments

	HCR-20 Rel.	HCR-20 SPJ I.V.	START V	START S	VRS	VRAG-R	PCL-R
HCR-20 Pres.	.92*** (33)	.67*** (33)	.92*** (33)	-.72*** (33)	.85*** (33)	.63*** (33)	.77*** (31)
HCR-20 Rel.		.62*** (33)	.82*** (33)	-.66*** (33)	.78*** (33)	.61*** (33)	.82*** (31)
HCR SPJ I.V.			.79*** (33)	-.48** (33)	.43*** (33)	.14 (33)	.26 (31)
START V				-.64*** (33)	.75*** (33)	.48** (33)	.59*** (31)
START S					-.51*** (33)	-.30* (33)	-.54*** (31)
VRS						.85*** (33)	.82*** (31)
VRAG-R							.80*** (31)

Note. *n* values vary by analysis and are provided in parentheses within cells containing under *r* values.
For the HCR-20, Pres. = Presence score, Rel. = Relevance score, SPJ I.V. = Structured Professional Judgement Summary Statement for Imminent Violence. For the START, V = Vulnerability score and S = Strength score.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 3.5

Intercorrelations among Dynamic and Static Risk Measure Components at Pretreatment

	HCR-20 C Pres.	HCR-20 R Pres.	HCR-20 H Rel.	HCR-20 C Rel.	HCR-20 R Rel.	START V	VRS Stat.	VRS Dyn.
HCR-20 H Pres.	.24 (33)	.51*** (33)	.93*** (33)	.32* (33)	.60*** (33)	.68*** (33)	.57*** (33)	.85*** (33)
HCR-20 C Pres.		.21 (34)	.16 (33)	.94*** (34)	.04 (34)	.82*** (33)	.12 (33)	.45** (33)
HCR-20 R Pres.			.31* (33)	.18 (34)	.76*** (34)	.44** (33)	.36* (33)	.59*** (33)
HCR-20 H Rel.				.26 (33)	.48** (33)	.54*** (33)	.46** (33)	.75*** (33)
HCR-20 C Rel.					.09 (34)	.84*** (33)	.13 (33)	.47** (33)
HCR-20 R Rel.						.34* (33)	.13 (33)	.56*** (33)
START V							.46** (33)	.76*** (33)
VRS Stat.								.55*** (33)

Note. n values vary by analysis and are provided in parentheses within cells containing under *r* values.

For the HCR-20, H = Historical scale, C = Clinical scale, R = Risk Management scale, Pres. = Presence scores, and Rel. = Relevance Scores. For the START, V = Vulnerability score. For the VRS, Stat. = Static subtotal and Dyn. = Dynamic subtotal.

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

Table 3.6

Measurement of Dynamic Changes in Risk from Pretreatment to Posttreatment

Instrument	Time 1 *M* (SD)	Time 2 *M* (SD)	Cohen's *d*	*t*
VRS				
Dynamic	42.37 (11.34)	44.45 (11.40)	.01	-1.02
Static	8.53 (3.49)	8.53 (3.49)	-	-
Total	50.82 (13.56)	50.89 (13.64)	.01	-1.00
HCR-20				
Presence				
C	6.13 (3.18)	6.13 (3.44)	.00	.00
R	8.47 (1.25)	8.47 (1.25)	-	-
Total	29.87 (7.07)	29.80 (7.13)	.01	.15
Relevance				
C	6.40 (3.00)	6.00 (3.36)	.13	.72
R	8.47 (1.19)	8.33 (1.11)	.12	1.00
Total	27.80 (7.19)	27.27 (7.05)	.07	.94
START				
Vulnerability	25.76 (8.75)	25.49 (9.23)	.03	.56
Strength	2.29 (2.40)	2.43 (2.70)	.05	-.50

Note. $n = 15$.

*$p < .05$, two-tailed. **$p < .01$, two-tailed. ***$p < .001$, two-tailed.

small for all instruments, with the HCR-20^{V3} Clinical and Risk Management Relevance scores producing the largest effects (d = .13 and d = .12).

Given the possibility of both reductions and increases in risk, the absolute values of changes were also computed (Table 3.7). A review of these descriptive statistics indicated that the mean difference in scores for the HCR-20^{V3} Clinical and Total scores, as well as the START Vulnerability score, were greater than 1 point. The mean absolute values of changes were less than 1 point for each of the VRS scores and the START Strength score.

3.3.4 Research Objective 2: Prediction of Inpatient Violence on Forensic Psychiatric Inpatient Units

In this study, the predictive validity of various instruments was assessed through ROC analyses (Tables 3.8, 3.9, and 3.10). Consistent with the results of Study 1, the magnitudes of AUCs are described here using Rice and Harris' (2005) guidelines, with .56 representing a small effect, .64 representing a medium effect, and .71 representing a large effect. It was hypothesized that total scores from each of the study instruments would significantly predict inpatient violence. Averaging across all study periods, the HCR-20^{V3} Presence and Relevance scores, VRS Total score, and START Vulnerability score each demonstrated medium sized AUC values, while the HCR-20^{V3} SPJ Imminent Violence summary rating produced a large AUC value. The following instruments did not demonstrate significant predictive validity for inpatient aggression: the START Strength score, the VRAG-R, and the PCL-R. Analyzing the three study periods separately, variability in effects was observed. Every measure that demonstrated significant predictive validity overall also demonstrated predictive validity for the first follow-up period; in contrast, only the HCR-20^{V3} Clinical scale scores, Imminent Violence SPJ summary statement, and START Vulnerability score demonstrated significant predictive validity in either of the second or third follow-up periods.

In addition to testing the hypotheses about instrument total scores, ROC analyses were conducted to explore the predictive validity of instrument components and subscales. With regard to the HCR-20^{V3}, results were consistent across Presence and Relevance scores: Historical scale scores produced small but significant AUC values, Clinical scale scores produced large and significant AUC values, and Risk Management scores did not produce significant AUC values. The VRS components both produced small and significant predictive effects. The various

Table 3.7

Descriptive Statistics: Absolute Values of Dynamic Changes in Risk from Pretreatment to Posttreatment

Instrument	*M* (SD)
HCR-20	
Presence	
Clinical	1.07 (1.33)
Total	1.13 (1.30)
Relevance	
Clinical	1.47 (1.60)
Total	1.60 (1.55)
START	
Vulnerability	1.33 (1.23)
Strength	.67 (.82)
VRS	
Dynamic	.08 (.29)
Static	.00 (.00)
Total	.07 (.28)

Note. $n = 15$.

Table 3.8

Receiver Operating Characteristic Analyses: Area Under the Curve Statistics for the HCR-20 and the Prediction of Inpatient Aggression

Instrument	Month(s)	*n*	AUC	95% CI
HCR-20 Total Pres.	1	302	.85***	(.76, .93)
	2	276	.59	(.48, .71)
	3	280	.59	(.40, .79)
	Combined	858	.68***	(.61, .76)
HCR-20 H Pres.	1	302	.79***	(.70, .88)
	2	276	.48	(.36, .60)
	3	280	.47	(.30, .64)
	Combined	858	.60*	(.52, .67)
HCR-20 C Pres.	1	302	.82***	(.74, .90)
	2	276	.70**	(.59, .82)
	3	308	.72***	(.61, .84)
	Combined	886	.74***	(.67, .80)
HCR-20 R Pres.	1	302	.57	(.46, .67)
	2	276	.44	(.30, .58)
	3	308	.51	(.32, .69)
	Combined	886	.53	(.44, .61)
HCR-20 Total Rel.	1	302	.87***	(.79, .95)
	2	276	.56	(.45, .67)
	3	280	.59	(.40, .77)
	Combined	858	.68***	(.59, .76)
HCR-20 H Rel.	1	302	.85***	(.76, .93)
	2	276	.47	(.35, .58)
	3	280	.48	(.30, .66)
	Combined	858	.61**	(.53, .69)
HCR-20 C Rel.	1	302	.82***	(.76, .89)
	2	276	.65*	(.54, .76)
	3	308	.72*	(.61, .84)
	Combined	886	.71***	(.65, .77)
HCR-20 R Rel.	1	302	.57	(.48, .65)
	2	276	.40	(.28, .51)
	3	308	.45	(.31, .60)
	Combined	886	.50	(.43, .58)
HCR-20 SPJ (CP)	1	302	.64*	(.55, .74)
	2	276	.55	(.43, .68)
	3	308	.55	(.39, .71)
	Combined	886	.58*	(.51, .65)
HCR-20 SPJ (SH)	1	302	.54	(.41, .67)
	2	276	.45	(.32, .58)
	3	308	.55	(.37, .72)
	Combined	886	.51	(.43, .78)
HCR-20 SPJ (IV)	1	302	.82***	(.74, .91)

2	276	.71**	(.61, .82)
3	308	.67†	(.53, .80)
Combined	886	.72*	(.66, .78)

Note. For the HCR-20, H = Historical scale, C = Clinical scale, R = Risk Management scale, Pres. = Presence score, Rel. = Relevance score, SPJ = Structured Professional Judgement, CP = Case Prioritization, SH = Severe Harm, IV = Imminent Violence.
†$p < .06$. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 3.9

Receiver Operating Characteristic Analyses: Area Under the Curve Statistics for the Prediction of Inpatient Aggression by Month

Instrument	Month(s)	*n*	AUC	95% CI
START Vulnerability	1	302	.80***	(.73, .86)
	2	276	.65*	(.56, .74)
	3	280	.66	(.54, .77)
	Combined	858	.70***	(.65, .75)
START Strength	1	302	.55	(.47, .63)
	2	276	.54	(.41, .67)
	3	280	.62	(.43, .80)
	Combined	858	.57	(.50, .64)
VRS Total	1	302	.81***	(.75, .88)
	2	276	.52	(.39, .65)
	3	280	.54	(.32, .76)
	Combined	858	.65***	(.57, .73)
VRS Dynamic	1	302	.79***	(.72, .85)
	2	276	.50	(.37, .62)
	3	280	.49	(.30, .68)
	Combined	858	.62**	(.55, .69)
VRS Static	1	302	.72***	(.62, .81)
	2	276	.62	(.49, .74)
	3	280	.55	(.40, .70)
	Combined	858	.63**	(.56, .70)
VRAG-R	1	302	.65*	(.55, .74)
	2	276	.51	(.37, .64)
	3	280	.46	(.27, .65)
	Combined	858	.54	(.47, .62)

Note. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 3.10

Receiver Operating Characteristic Analyses: Area Under the Curve Statistics for the PCL-R and the Prediction of Inpatient Aggression

Instrument	Assessment Period	*n*	AUC	95% CI
PCL-R Total	1	269	.73***	(.66, .80)
	2	276	.43	(.32, .54)
	3	280	.44	(.29, .58)
	Combined	825	.55	(.48, .62)
Facet 1	1	269	.72***	(.65, .80)
	2	276	.38	(.26, .50)
	3	280	.35	(.21, .49)
	Combined	825	.50	(.43, .58)
Facet 2	1	241	.50	(.43, .57)
	2	248	.46	(.33, .59)
	3	252	.50	(.31, .69)
	Combined	741	.51	(.45, .58)
Facet 3	1	269	.80***	(.71, .89)
	2	276	.44	(.32, .56)
	3	280	.56	(.40, .72)
	Combined	825	.61**	(.53, .69)
Facet 4	1	274	.74***	(.66, .83)
	2	276	.47	(.38, .57)
	3	280	.39	(.26, .51)
	Combined	830	.57	(.50, .64)
Factor 1	1	241	.71**	(.65, .77)
	2	248	.42	(.31, .53)
	3	252	.39	(.26, .52)
	Combined	741	.54	(.47, .60)
Factor 2	1	269	.82***	(.75, .89)
	2	276	.46	(.35, .56)
	3	280	.46	(.30, .62)
	Combined	825	.61*	(.53, .68)

Note. *$p < .05$. **$p < .01$. ***$p < .001$.

components of the PCL-R were analyzed, and only Facet 3 and Factor 2 produced small and significant AUCs.

3.3.5 Research Objective 3: Contribution of Dynamic Risk Measures to the Prediction of Inpatient Violence

Logistic regression analyses were conducted to assess whether the dynamic components of the HCR-20^{V3} and VRS would demonstrate incremental predictive validity for inpatient violence, controlling for static factors. Overall models, combining the respective static and dynamic components, produced significant χ^2 statistics for the prediction of inpatient aggression (Table 3.11). HCR-20^{V3} Clinical Presence and Relevance scores each demonstrated incremental predictive validity over Historical scale scores (Presence e^B = 1.42 and Relevance e^B = 1.39), and over a combination of Historical scale and Risk Management scores (Presence e^B = 1.41 and Relevance e^B = 1.38), for the prediction of inpatient violence. VRS Dynamic scores also demonstrated incremental prediction beyond Static scores (e^B = 1.04).

Similar logistic regression analyses were completed to assess the incremental predictive validity of dynamic components beyond a consistent measure of static risk, the VRAG-R (Table 3.12). All of the overall models, which combined a dynamic measure with the VRAG-R, significantly predicted inpatient violence. HCR-20^{V3} Clinical Presence and Relevance scores both demonstrated incremental prediction beyond the VRAG-R (Presence e^B = 1.45 and Relevance e^B = 1.42). START Vulnerability and Strength scores also significantly predicted inpatient violence controlling for the VRAG-R (Vulnerability e^B = 1.12 and Strength e^B = .81). Finally, the VRS Dynamic score also demonstrated incremental prediction beyond the VRAG-R (e^B = 1.10).

3.4 Discussion

3.4.1 Research Objective 1: Detecting Dynamic Risk: The VRS, START, and HCR-20^{V3}

To assess the extent to which the instruments detected dynamic changes in risk between assessment periods (i.e., over 28 day periods), a number of *t*-tests were conducted. None of the instruments were found to detect significant changes in risk, with Cohen's *d* values ranging from .00 to .13. These results are markedly different from the findings of Study 1, in which each of the dynamic measures demonstrated significant results for the detection of change. These discrepant findings were likely influenced by differences in design between the two studies. For example, Study 1 change analyses were based on differences between scores at admission and

Table 3.11

Logistic Regression Analyses: Incremental Contributions of Dynamic Variables over Static Variables for the Prediction of Inpatient Violence with HCR-20 and VRS

Predictor	*B*	*SE*	Wald	e^B	95% CI
HCR-20 Presence (*n* = 858)					
Historical	.06	.05	1.67	1.06	(.97, 1.16)
Clinical	.35	.08	21.34***	1.42	(1.22, 1.64)
Constant	-6.41	.95	45.57***	.00	
Model χ^2 (2) = 35.31***					
Historical	.07	.05	1.67	1.07	(.97, 1.19)
Clinical	.34	.08	20.64***	1.41	(1.22, 1.64)
Risk Management	-.05	.14	.13	.95	(.72, 1.25)
Constant	-6.08	1.31	21.51***		
Model χ^2 (3) = 35.44***					
HCR-20 Relevance (*n* = 858)					
Historical	.08	.05	2.43	1.09	(.98, 1.20)
Clinical	.33	.08	19.47***	1.39	(1.20, 1.61)
Constant	-6.47	.94	47.92***	.00	
Model χ^2 (2) = 32.90***					
Historical	.13	.07	3.32	1.13	(.99, 1.30)
Clinical	.32	.07	18.98***	1.38	(1.19, 1.59)
Risk Management	-.20	.18	1.27	.82	(.58, 1.16)
Constant	-5.38	1.32	16.73***	.01	
Historical					
Model χ^2 (3) = 34.19***					
VRS (*n* = 858)					
Static	.06	.04	2.18	1.07	(.98, 1.16)
Dynamic	.04	.02	4.11*	1.04	(1.00, 1.08)
Constant	-5.17	.83	38.73***	.006	
Model χ^2 (2) = 11.97**					

Note. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 3.12

Logistic Regression Analyses of the Relative Contributions of the VRAG-R and the Dynamic Components of the HCR-20, START, and VRS to the Prediction of Inpatient Violence

Model	*B*	*SE*	Wald	e^B	95% CI
1 *HCR-20 Presence* (*n* = 858)					
VRAG-R	.02	.01	2.38	1.02	(1.00, 1.04)
Clinical	.37	.08	23.34***	1.45	(1.25, 1.69)
Constant	-5.89	.72	67.26***	.00	
Model χ^2 (2) = 36.10***					
2 *HCR-20 Relevance* (*n* = 858)					
VRAG-R	.01	.01	.90	1.01	(.99, 1.03)
Clinical	.35	.08	21.55***	1.42	(1.22, 1.64)
Constant	-5.63	.67	69.65***	.00	
Model χ^2 (2) = 31.23***					
3 *START* (*n* = 858)					
VRAG-R	-.01	.01	.29	1.00	(.98, 1.01)
Vulnerability	.11	.03	17.08***	1.12	(1.06, 1.18)
Constant	-5.91	.80	55.14***	.00	
Model χ^2 (2) = 22.07***					
4 *START* (*n* = 858)					
VRAG-R	.00	.01	.16	1.00	(.99, 1.02)
Strength	-.21	.10	4.78*	.81	(.67, .98)
Constant	-2.50	.26	91.06***	.08	
Model χ^2 (2) = 7.62*					
5 *VRS* (*n* = 858)					
VRAG-R	-.03	.01	3.81*	.97	(.95, 1.00)
Dynamic	.09	.03	10.84***	1.10	(1.04, 1.16)
Constant	-6.46	1.12	33.32	.00	
Model χ^2 (2) = 13.67***					

Note. *$p < .05$. **$p < .01$. ***$p < .001$.

discharge/transfer, and discharges/transfers were contingent upon reductions in risk or acute psychiatric symptoms. In contrast, Study 2 change analyses were simply based on differences observed at two time points, among patients who by necessity, had not met criteria for a discharge or transfer. Furthermore, the current study excluded a number of individuals whose dates of admission did not coincide with the common admission and follow-up periods. As a result, patients who were briefly admitted for a few weeks for the stabilization of acute mental health symptoms could easily have been excluded from any change analyses, despite being present during a portion of this research. Thus, the absence of significant change findings was likely influenced by design characteristics of the study, which differentially included patients demonstrating acutely elevated risk factors, rather than by instrument characteristics alone. Indeed, a review of the absolute values of change scores indicated that each of the HCR-20^{V3} Total and Clinical scores, and the START Vulnerability score, demonstrated a mean change score greater than one point (i.e., either an increase or decrease in score).

3.4.2 Research Objective 2: Prediction of Inpatient Violence on Forensic Psychiatric Inpatient Units

Predictive validity analyses for this study looked exclusively at inpatient outcomes. Instruments specifically designed to assess clinical/dynamic risk factors for violence, including the HCR-20^{V3}, VRS, and the START, each demonstrated predictive validity for the prediction of inpatient violence; instruments that were not designed for this purpose, including the VRAG-R and the PCL-R, did not. Certain components of individual tools that were intended to capture relatively static risk factors, such as the HCR-20^{V3} Historical scale scores and the VRS Static score, did predict inpatient violence in this study, but it was the dynamic subscales that performed best, with the HCR-20^{V3} Clinical scale scores producing the largest effects.

Both similarities and differences were noted between the current findings and the limited previous research in this area using the clinical/dynamic tools. Results with the VRS for example, were slightly lower than, but also similar to the findings of Dolan and Fullam (2007) and Dolan and colleagues (2008), with the current AUC 95% confidence interval including both of their reported AUCs. Results of analyses using the HCR-20 Version 3's two total scores (Presence and Relevance AUCs = .68) and the Imminent Violence SPJ summary statement (AUC = .72) were quite consistent with meta-analytic findings using the previous versions of the tool (AUC = .69; Hogan & Ennis, 2010); the AUC for the Case Prioritization SPJ summary

statement was somewhat lower (AUC = .58), but it is noted that this finding likely does not reflect an accurate assessment of the predictive validity of the measure, given that there was minimal variability in this sample (total ratings of: Low = 0, Moderate = 6, High = 29). HCR-20 Version 3 Clinical scale scores (Presence AUC = .74 and Relevance AUC = .71) demonstrated strong predictive validity in this study, and effects were also consistent with previous research using the previous iterations' Clinical scale (AUC = .70; Hogan & Ennis, 2010).

The current findings pertaining to the START Vulnerability scores supported its predictive validity for inpatient aggression, and produced an AUC 95% confidence interval that overlapped with those of similar studies (Chu et al., 2011; Desmarais et al., 2012; Wilson et al., 2013). START Strength scores did not demonstrate significant predictive validity in this study, and while the AUC 95% confidence interval overlapped with the results of Chu and colleagues (2011), it did not overlap with the results of Desmarais and colleagues (2012) or Wilson and colleagues (2013). The observed discrepancy between the current results and those of previous research was likely influenced by limited variability in scores, resulting from the relatively low Strength scores observed in this sample overall (mean = 2.53, SD = 3.00). Indeed, previous research samples producing stronger predictive effects demonstrated considerably higher Strength scores overall and greater variability, including Desmarais and colleagues' (2012) sample (mean = 18.46, SD = 8.08) and Wilson and colleagues' (2013) sample (mean = 18.30, SD = 7.52).

Both of the instruments that were not designed specifically to capture dynamic risk factors for violence failed to demonstrate predictive validity for inpatient aggression in this study. While the VRAG-R itself has not yet been studied extensively with regard to the prediction of inpatient violence, the absence of predictive validity was notably inconsistent with previous research using the previous version of the VRAG (Doyle et al., 2002; Vitacco et al., 2012b). The result was however consistent with the results of Study 1 from the current program of research. This being said, the follow-up period in the current study was notably shorter than those employed by Doyle and colleagues (2002) and Vitacco and colleagues (2012b), which ranged from approximately 3 to 12 months. Given that the VRAG-R was developed to assess long-term risk for violent recidivism, it is certainly possible that in keeping with static risk factors more generally, the instrument simply performs better over longer follow-up periods - additional research will be necessary to clarify this point.

Like the VRAG-R, the PCL-R did not significantly predict inpatient aggression in this study. The PCL-R AUC 95% confidence interval overlapped with the 95% credibility interval (credibility interval is a Bayesian analogue to a confidence interval) reported in Hogan and Ennis' (2010) meta-analysis, although the overall effect observed in the current study does fall on the low end of the range of previous effects. Observed differences between previous findings with the PCL-R and the current study may also have been influenced by the follow-up period in the current research, and the instrument's greater relevance to long-term risk. The PCL-R was designed to capture stable characteristics, rather than evaluate short-term risk, and previous studies producing stronger results for the prediction of inpatient violence have used longer follow-up periods (Gray et al., 2003; Reimann, 2008).

The results discussed thus far were produced by analyzing the entire sample as a whole, but there were also interesting findings from analyses breaking the data into distinct assessment and follow-up periods. Most notably, considerable differences in predictive effects were observed among separate follow-up periods for the same instruments. For example, HCR-20^{V3} Presence total scores and VRS total scores produced non-overlapping 95% confidence intervals for their respective AUCs at months 1 and 2. Other measures, like the START Vulnerability score, produced overlapping 95% confidence intervals for AUCs at each monthly assessment (minimally, in the case of START Vulnerability scores), but AUCs only reached significance in particular months. The only measures that were consistent and significant predictors of inpatient aggression across the monthly follow-up periods were HCR-20^{V3} Clinical scale Presence and Relevance scores. HCR-20 Clinical scales scores have previously been found to be among the strongest predictors of inpatient aggression (Hogan & Ennis, 2010), and in this study they also provided more consistent utility than other measures. While the data available in this study do not shed light on potential moderators of the predictive effects produced by the other measures, the results nonetheless suggest that HCR-20 Clinical scale scores may be more consistent predictors of aggression than the other study measures, when used as a monthly screening tool in forensic psychiatric settings.

3.4.3 Research Objective 3: Contribution of Dynamic Risk Measures to the Prediction of Inpatient Violence

Logistic regression analyses were used to assess the relative contributions of different types of risk measures to the assessment of risk for inpatient violence. Dynamic risk measures

incrementally predicted inpatient aggression in this study beyond static measures, regardless of the model used. For example, HCR-20^{V3} Clinical scale scores (Presence and Relevance) incrementally predicted aggression, controlling for the respective Historical scales, and controlling for both Historical scale scores and Risk Management scale scores. Keeping the other measures constant, one point increases in the various Clinical scores were associated with increases in the odds of aggression of between 38% and 42%. A similar result was found for a model combining the VRS Dynamic and Static scores; holding Static score constant, an increase in Dynamic score of one point was associated with an increase of 4% in the odds of inpatient violence. This general pattern was also observed for models combining the various dynamic risk scores, taken from the VRS, HCR-20^{V3}, and START, with a common measure of static risk, the VRAG-R.

The incremental validity of dynamic measures observed in this study was generally consistent with the existing literature. Desmarais and colleagues (2012) for example, demonstrated that START Vulnerability scores were incrementally predictive of inpatient aggression beyond HCR-20 Historical scale scores. Wilson and colleagues (2013) found that a combination of the HCR-20 Clinical and Risk Management scales was incrementally predictive beyond the Historical scale, and that the START Vulnerability score was also incrementally predictive beyond the HCR-20 Historical scale. Using null hypothesis significance testing, Vitacco and colleagues (2012b) did not find that the HCR-20 Clinical scale incrementally predicted inpatient aggression beyond a combination of the Historical scale, PCL-R, and VRAG; however, they did find that the Clinical scale produced a large odds ratio (e^B = 1.36), and their overall results were likely influenced by the use of multiple intercorrelated measures to control for static risk (described in the Discussion section of Study 1). Thus, the current findings have contributed further support for the use of risk instruments including dynamic or clinical risk factors.

3.4.4 Strengths and Limitations

In addition to the preceding discussion points, some strengths of the study are worth noting. The prospective design increased the strength of conclusions that may be drawn from the predictive validity analyses. Additionally, the combination of both between and within subjects comparisons allowed for an indirect assessment of the dynamic nature of the predictors, broadly following Wilson and colleagues' (2013) example. Consistent with Study 1, this research made

a valuable contribution to the limited literature on risk assessment for forensic psychiatric inpatient violence, and represents one of the first attempts to validate two new instruments, the VRAG-R and the HCR-20^{V3}, for the assessment of forensic psychiatric inpatient violence in Canada.

Potential limitations were also noted. Like Study 1, START analyses were based exclusively on summed total scores, rather than the SPJ summary statements the authors recommend for clinical practice. The use of file data to code study instruments meant that the risk assessments were limited by the quality of the data available in institutional files, with the result that certain instruments could not be scored for particular patients. This being said, it should be noted that this may also reflect a limitation of the instruments themselves when applied to inpatient violence. To explain further, given that clinicians would require multiple sources of information (i.e., more than interview data alone) to corroborate conclusions and conduct a valid risk assessment (American Psychological Association, 2013), and given that some of the participants who were admitted because of acute mental health crises were not capable/willing to engage in an interview upon admission, the lack of documentation available in the files would likely have limited the scoring of instruments regardless of the methods employed. Another limitation of this study was the fact that the small number of patients present across multiple the study periods precluded the direct assessment of the dynamic nature of the risk measures. Finally, while treating multiple assessments of the same individuals as separate units of analysis was consistent with the existing literature on inpatient and institutional violence (Almvik et al., 2000; Chu et al., 2013; Wilson et al., 2010; Wilson et al., 2013), and allowed for useful information to the gleaned within the practical constraints of the study, it nonetheless technically violates the statistical assumption of independence of observations.

3.5 Conclusions

This study demonstrated that a number of existing structured violence risk assessment instruments were capable of assessing risk for forensic psychiatric inpatient violence. In particular, the instruments that were designed to include clinical or dynamic risk factors produced the strongest results, when compared to instruments that were not designed for this purpose. Clinical and dynamic variables were the most consistent predictors over multiple time periods, and they demonstrated incremental predictive validity, controlling for static risk factors. These results indicate that professionals interested in assessing risk for this outcome have a

number of potentially useful tools available to them, including the latest version of the HCR-20, and that they would be well advised to consider a measure of clinical or dynamic risk as part of assessment procedures.

Chapter 4: GENERAL DISCUSSION

4.1 Overview

The current program of research was designed to assess the predictive validity of existing violence risk assessment technology within forensic psychiatric institutions. More specifically, the primary goal was to evaluate the utility of a number of new and existing instruments for the assessment of risk for inpatient violence. While the literature in this area is now growing, this research was intended to make a contribution to an area that is still in need of further study. With the limited resources available to forensic mental health programs in mind, this research was specifically designed to evaluate whether tools designed and validated for other purposes (e.g., community recidivism) could also be applied to inpatient violence risk assessment. In addition to the primary focus on assessing risk for inpatient violence, this research also evaluated other aspects of violence risk assessment among forensic psychiatric inpatients. Additional foci included the dynamic nature of violence risk and the relative contributions of different types of risk factors to the assessment of institutional outcomes.

Participants in the two studies that comprise this research program were forensic psychiatric inpatients, who were assessed either retrospectively (n = 99) or prospectively (n = 19) for their risk for violence, based on information available in institutional files. Approximately half of the patients in Study 1 (48%) and three quarters of the participants in Study 2 had been found NCR for their index offences; the remaining participants were admitted for forensic assessments, acute mental health care, or for treatment/supervision subsequent to having been found unfit to stand trial. Generally speaking, the samples were broadly similar to Latimer and Lawrence's (2006) description of the forensic mental health population in Canada, based on a variety of demographic variables (e.g., gender), psychiatric diagnoses (e.g., rates of schizophrenia and related disorders), and offence characteristics (e.g., rates of violent index offences). The two samples differed somewhat in risk profiles, as discussed previously, with the second sample scoring higher on the PCL-R and measures of risk. Nonetheless, it is likely that the current results are broadly applicable to other typical forensic psychiatric settings in Canada.

4.2 Research Objective 1: Detecting Dynamic Risk: The VRS, the START, and the HCR-20^{V3}

With regard to the detection of dynamic changes in risk, the study measures' performance varied between studies. When used to measure changes in risk factors occurring over a longer

term in Study 1 (mean follow-up = 19.37 months, SD = 26.97 months), the dynamic risk instruments performed well for this purpose; when used to measure changes over shorter, monthly follow-ups, none of the instruments captured significant changes in risk overall. Part of this difference was likely influenced by the different paradigms employed in the studies, and the method employed to analyze differences, rather than simply duration of follow-up. Given that admissions to the maximum security unit were contingent upon perceived risk, and given that discharges/transfers were generally contingent upon perceived reductions in risk, it follows that a reduction in risk scores was expected, making paired-samples *t*-tests appropriate analytic tools. In Study 2 on the other hand, participants were only re-scored on the measures if they were not discharged or transferred, and thus it can be inferred that the clinicians working with the patients had not yet observed sufficient reductions in risk to facilitate such a move. Therefore, the distribution of scores at time 2 would not necessarily be expected to be lower than the distribution observed at time 1, and thus the paired-samples *t*-tests may not have reflected the instruments' utility for detecting changes in risk (i.e., increases or decreases). For some of the measures, absolute values of change scores indicated an average change of more than one point, with both increases and decreases included. Overall, the dynamic study instruments demonstrated efficacy for the detection of dynamic changes in risk, at least over the longer term, among forensic psychiatric inpatients.

4.3 Research Objective 2: Assessment of Risk for Inpatient Violence on Forensic Psychiatric Inpatient Units

Through various analyses, a number of the study instruments demonstrated predictive validity for forensic psychiatric inpatient violence, as operationalized in this program of research. It is important to recall, when interpreting the results of the predictive validity analyses, that the relatively broad outcome variable included serious acts of aggression that did not necessarily result in physical injury or harm to an individual (e.g., attempted but not completed acts of violence or significantly threatening behaviours).

In keeping with Daffern's (2007) discussion points, dynamic and clinical measures demonstrated superior predictive validity when compared to static measures. Across both studies, more specialized measures designed to capture relevant dynamic and clinical variables, such as the HCR-20^{V3} Clinical scale, HCR-20^{V3} Imminent Violence SPJ summary statement, and the START Vulnerability score, produced the largest predictive effects, while measures that were

not designed for this purpose, such as the VRAG-R and PCL-R Total score, produced small and insignificant predictive effects. Interestingly, the VRS, which was designed to capture dynamic risk factors that are arguably more stable than those captured by the HCR-20^{V3} Clinical scale or the START Vulnerability score, produced predictive effects that were smaller than the other dynamic instruments, but larger than those produced by the VRAG-R and PCL-R. Thus, the predictive validity of the measures used in this program appeared to roughly line up with a conceptual continuum, ranging from purpose-built dynamic measures to measures designed for other purposes.

When considering the application of these findings within active forensic mental health programs, it is important to note a few points about the time frames of inpatient violence assessments. For instance, it should be noted that the results of the first study applied to variable follow-ups; they do not shed light on the probability of violence over a fixed interval of interest. The results of the second study on the other hand, may be interpreted as reflecting the relationship between scores on the instruments and the probability of a violent incident on a single day, at any time up to four weeks after an assessment. It is important to keep these considerations in mind, because different risk factors can be expected to apply differently depending on the length of follow up the assessor wishes to address (e.g., 24 hours or 12 months). Furthermore, the integration of information about long-term risk gathered through static risk variables (e.g., VRS Static score) and about imminent risk gathered through clinical/dynamic risk variables, remains a complex task.

4.4 Research Objective 3: Contribution of Dynamic Risk Measures to the Prediction of Inpatient Violence

The observed trends of dynamic measures demonstrating the strongest predictive validity for inpatient violence were also reflected in analyses directly comparing the different measures. Across both studies, the respective clinical/dynamic components of the HCR-20^{V3} and VRS demonstrated incremental prediction beyond the historical/static components. Similarly, the HCR-20^{V3} Clinical scale and START Vulnerability and Strength scores demonstrated incremental predictive validity beyond the VRAG-R for inpatient violence across both studies. Unfortunately, the predictive validity of measured changes in risk for the prediction of inpatient violence could not be assessed directly in either study. However, taken together, the findings of Study 1 that indicated that the ostensibly dynamic instruments can detect changes, and the

findings across both studies supporting the predictive validity of these measures, did provide some indirect evidence for the relevance of dynamic risk factors.

4.5 Research Objective 4: The Relationship between Dynamic Changes in Risk and Inpatient Violence

Research Objective 4 was assessed only through Study 1. The analyses of the concurrent validity of dynamic change scores and inpatient violence produced some interesting results. Each of the dynamic instruments produced change scores that demonstrated incremental concurrent validity with decreased inpatient violence, beyond pretreatment scores. These findings provided at least partial support for the dynamic nature of the risk factors. Additionally, the HCR-20^{V3}'s Relevance scores demonstrated more significant relationships with inpatient violence than did Presence scores. This finding provides some support for the contention that because Relevance scores involve a higher threshold for change than do Presence scores, corresponding changes are likely to reflect more stable, risk-relevant changes.

4.6 Evidence for the Validity of New Instruments: The HCR-20^{V3} and VRAG-R

The current program of research generally supported the use of the HCR-20^{V3} for the assessment of risk for inpatient violence among forensic psychiatric inpatients in Canada. Data from Study 1 provided support for the interrater reliability of the instrument, and intercorrelations observed between the HCR-20^{V3} and other instruments in both studies demonstrated evidence for convergent validity. Predictive validity analyses demonstrated that the various HCR-20^{V3} scores and SPJ summary statements were strong and significant predictors of inpatient violence, with Clinical scores outperforming Historical scores for inpatient outcomes.

The VRAG-R was not a significant predictor of inpatient aggression in this study. One perceived clinical limitation of the VRAG-R noted during the data collection phases of this research program, that may also have impacted the current findings, was the potential impact that variations in data may have on "at least" decision making (i.e., in the absence of complete data, choosing the highest confirmed option) and the prorating of scores. An unknown number of cases required "at least" decision making, but 36% of the VRAG-R total scores in Study 1 and 44% of those in Study 2 required prorating. While these score adjustments were consistent with VRAG-R scoring instructions, they nonetheless could have significantly influenced the final scores. Clinicians using the VRAG-R scores for clinical purposes should seriously consider the

potential impact of incomplete data on their scores for a particular individual. Discretion is strongly advised when reporting individual results, given that the omission or addition of a single point on the VRAG-R may push a given score into a neighbouring risk category.

4.7 Future Directions

While this research has provided some support for the use of existing instruments to assess risk for forensic psychiatric inpatient violence, it has also highlighted areas in need of further study. For example, existing research has validated particular risk instruments for follow-up periods ranging from a single day (Chu et al., 2013) up to a full year (Desmarais et al., 2012), but it is unclear how to best synthesize and integrate these findings for clinical applications. Questions remain to be answered about what combination of risk factors, reassessed at what intervals, represent best practice for the assessment of risk for inpatient violence. Further studies may look to clarify the relative and incremental contributions of daily, weekly, or monthly ratings of risk measures capturing risk factors falling on the continuum ranging from the acute (e.g., Brøset Violence Checklist) to the static (e.g., HCR-20^{V3} Historical scale, VRS Static score). It is also worth directly assessing dynamic changes in risk and their incremental predictive validity for inpatient violence, as this has generally been assessed only indirectly in the current studies and in previous research (Wilson et al., 2013).

Another potential direction for future study involves the integration of assessment and treatment research in forensic mental health. As Serin and colleagues' (2013) pointed out in their review of the research on offender change and recidivism, the adequacy of contemporary conceptualizations and measurement of change remains to be explored. The current research examined instruments that capture dynamic changes in risk through diverse methods, including: relatively straightforward summing of changes in presence of items, as measured by the START; progression through the stages of a modified version of Prochaska and colleagues' (1992) TTM, as measured by the VRS; and changes in the Relevance of risk factors, as measured and defined by the HCR-20^{V3}. The current research demonstrated incremental relationships between change scores and inpatient violence for each of the dynamic measures, but further research is necessary to directly investigate the predictive validity and utility of different models for the assessment of dynamic risk and treatment-related change.

As research on risk assessment practices establishes the credibility of dynamic and potentially causal (Andrews & Bonta, 2010; Kraemer et al., 1997) risk factors, it may also

provide feedback to evaluators and developers of forensic treatment programs. According to Lipsey, Landenberger, and Wilson (2007), cognitive-behavioral interventions were among the most empirically supported methods of offender rehabilitation, and they tended to focus on factors like procriminal cognitions, anger management skills, and personal accountability. Serin and colleagues (2013) reported that the majority of studies of change among violent offenders in their review examined measures of constructs like negative affect, anger, or impulsivity. A cursory review of the items that comprise the risk instruments examined in the current research program indicates that a number of the validated risk factors are being targeted only indirectly, if at all, by these types of interventions. Indeed, if one were to use the risk measures themselves to inform the identification of further targets for intervention, one might see that other important criminogenic needs, such as educational/employment activities, could be targeted through behaviorally-oriented approaches, like contingency management programs (Gendreau, Listwan, Kuhns, & Exum, 2014). Furthermore, by integrating these risk assessment measures more closely into treatment programs, treatment providers may find it easier to facilitate ongoing outcome monitoring and to move from a one-size-fits-all treatment approach to one that is more focused on the individual's risk formulation (Hart et al., 2011). The VRS was designed specifically to guide interventions (Wong & Gordon, 2006) in this manner, but other instruments capturing dynamic risk factors, such as the HCR-20^{V3}, may perform a similar function and inform more effective rehabilitation efforts.

4.8 Conclusions

Despite being associated with serious costs and negative consequences, forensic psychiatric inpatient violence has received relatively little research attention. Identifying reliable and valid assessment tools is a necessary first step towards applying established principles of offender programming (RNR; Andrews & Bonta, 2010; Bonta & Andrews, 2007) towards the reduction of aggression within forensic mental health settings. The current findings provided further evidence that valid violence risk assessments are possible within these settings, using existing and newly developed/revised structured instruments. Results further indicated that dynamic or clinical variables represent an important piece of valid assessments of risk for inpatient violence. Additional research will be necessary to establish which predictors are best suited to particular time periods, outcomes, and settings. Nonetheless, it is hoped that the current results will assist and inform forensic mental health professionals' selection of instruments, and

facilitate improved distribution of resources and selection of interventions, towards the mitigation of violence in forensic hospitals.

References

Ahmed, A. G., Adamson, D., & Laprade, K. (2004). *Brockville Risk Checklist (3rd version): Interdisciplinary clinical risk assessment and management tool*. Brockville, ON: Brockville Mental Health Centre.

Almvik, R., Woods, P., & Rasmussen, K. (2000). The Brøset Violence Checklist (BVC): Sensitivity, specificity and inter-rater reliability. *Journal of Interpersonal Violence, 15,* 1284-1296. doi: 10.1177/088626000015012003

American Psychological Association. (2013). Specialty guidelines for forensic psychologists. *American Psychologist, 68,* 7-19.

Amore, M., Menchetti, M., Tonti, C., Scarlatti, F., Lundgren, E., Esposito, W., & Berardi, D. (2008). Predictors of violent behavior among acute psychiatric patients: Clinical study. *Psychiatry and Clinical Neurosciences, 62*(3), 247-255.

Andrews, D. A. (1982). *The Level of Supervision Inventory (LSI)*. Toronto, ON: Ministry of Correctional Services.

Andrews, D. A. & Bonta, J. (1995). *The Level of Service Inventory-Revised.* Toronto, ON: Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law, 16,* 39-55.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17,* 19-52.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *The Level of Service/Case Management Inventory (LS/CMI)*. Toronto, ON: Multi-Health Systems.

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

Andrews, D. A., Bonta, J., & Wormith, J. S. (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.

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87,* 84-94.

Bidna, H. (1975). Effects of increased security on prison violence. *Journal of Criminal Justice, 3,* 33-46.

Bjørkly, S., Eidhammer, G., & Selmer, L. E. (2014). Concurrent validity and clinical utility of the HCR-20^{V3} compared to the HCR-20: Similar tools but improved method. *Journal of Forensic Nursing, 10,* 234-242.

Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, CA: Sage.

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and rehabilitation* (Corrections Research User Report No. 2007-06). Ottawa, ON: Public Safety Canada.

Bonta, J., Harman, W. G., Hann, R. G., & Cormier, R. B. (1996). The prediction of recidivism among federally sentenced offenders: A revalidation of the SIR scale. *Canadian Journal of Criminology, 38,* 61–79.

Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin, 123*, 123-142.

Borum, R., Bartel, P., & Forth, A. (2002). *SAVRY: Manual for the Structured Assessment of Violence Risk in Youth.* Tampa, FL: Florida Mental Health Institute, University of South Florida.

Bowers, L., Allan, T., Simpson, A., Nijman, H., & Warren, J. (2007). Adverse incidents, patient flow and nursing workforce variables on acute psychiatric wards: The Tompkins Acute Ward Study. *International Journal of Social Psychiatry, 53,* 75-84.

Braithwaite, E., Charette, Y., Crocker, A. G., & 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*(4), 271-281.

Burgess, E. W. (1928). Factors determining success or failure on parole. In A. A. Bruce, A. J. Harno, E. W. Burgess, & J. Landesco (Eds.), *The workings of the indeterminate sentence law and the parole system in Illinois* (pp. 221–234). Springfield, IL: State Board of Parole.

Campbell, M. A., 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.

Carmel, H., & Hunter, M. (1989). Staff injuries from inpatient violence. *Hospital and Community Psychiatry, 40*, 41-46.

Chagigiorgis, H., Michel, S. F., Seto, M. C., Laprade, K., & Ahmed, A. G. (2013). Assessing short-term, dynamic changes in risk: The predictive validity of the Brockville Risk Checklist. *International Journal of Forensic Mental Health, 12*(4), 274-286.

Chu, C. M., Daffern, M., & Ogloff, J. R. P. (2013). Predicting aggression in acute inpatient psychiatric setting using BVC, DASA, and HCR-20 Clinical scale. *The Journal of Forensic Psychiatry & Psychology, 24*(2), 269-285.

Chu, C. M., Thomas, S. D. M., Ogloff, J. R. P., & 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*(4), 337-345.

Copas, J., & Marshall, P. (1998). The Offender Group Reconviction Scale: The statistical reconviction score for use by probation officers. *Journal of the Royal Statistical Society: Series C, 47*, 159–171.

Cornaggia, C. M., Beghi, M., Pavone, F., & Barale, F. (2011). Aggression in psychiatry wards: A systematic review. *Psychiatry Research, 189,* 10-20.

Corrigan, P. W., & Watson, A. C. (2005). Findings from the National Comorbidity Survey on the frequency of violent behavior in individuals with psychiatric disorders. *Psychiatry Research, 136,* 153-162.

Crocker, A. G., Seto, M. C., Nicholls, T. L., and Cote, G. (2013). Description and processing of individuals found Not Criminally Responsible on Account of Mental Disorder accused of "serious violent offences". Retrieved from https://ntp-ptn.org/NCRMD-SVO-NTPteam_March_2013.pdf

Cyr, J. J., & Paradis, J. (2012). The forensic float nurse: A new concept in the effective management of service delivery in a forensic program. *Journal of Forensic Nursing, 8,* 188-194.

Daffern, M. (2007). The predictive validity and practical utility of structured schemes used to assess risk for aggression in psychiatric inpatient settings. *Aggression and Violent Behavior, 12,* 116-130.

Daffern, M., & Howells, K. (2002). Psychiatric inpatient aggression: A review of structural and functional assessment approaches. *Aggression and Violent Behavior, 7,* 477–497.

Daffern, M., Howells, K., Ogloff, J., & Lee, J. (2005). Individual characteristics predisposing patients to aggression in a forensic psychiatric hospital. *The Journal of Forensic Psychiatry & Psychology, 16,* 729-746.

Dauvergne, M. (2012). Adult correctional statistics in Canada, 2010/2011. *Juristat*. Retrieved from http:// www.statcan.gc.ca

de Borba Telles, L. E., Folino, J. O., & Taborda, J. G. V. (2011). Incidence of violent and antisocial behavior in a forensic psychiatric population. *Trends in Psychiatry and Psychotherapy, 33,* 3-7.

de Vogel, V., van den Broek, E., & de Vries Robbé, M. (2014). The value of the HCR-20^{V3} in Dutch forensic psychiatric patients. *International Journal of Forensic Mental Health, 13*, 109–121.

de Vries Robbé, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. I. (2015). Changes in dynamic risk and protective factors for violence during inpatient forensic psychiatric treatment: Predicting reductions in postdischarge community recidivism. *Law and Human Behavior, 39,* 53-61.

Desmarais, S. L., Nicholls, T. L., Wilson, C. M., & Brink, J. (2012). Using dynamic risk and protective factors to predict inpatient aggression: Reliability and validity of START assessments. *Psychological Assessment, 24*, 685-700.

Dolan, M. & Davies, G. (2006). Psychopathy and institutional outcome in patients with schizophrenia in forensic settings in the UK. *Schizophrenia Research, 81,* 277-281.

Dolan, M., & Fullam, R. (2007). The validity of the Violence Risk Scale Second Edition (VRS-2) in a British forensic inpatient sample. *The Journal of Forensic Psychiatric & Psychology, 18*(3), 381-393.

Dolan, M., Fullam, R., Logan, C., & Davies, G. (2008). The Violence Risk Scale Second Edition (VRS-2) as a predictor of institutional violence in a British forensic inpatient sample. *Psychiatry Research, 158*, 55-65.

Douglas, K. S. (2014). Version 3 of the Historical-Clinical-Risk Management-20 (HCR-20^{V3}): Relevance to violence risk assessment and management in forensic conditional release contexts. *Behavioral Sciences & the Law, 32*(5), 557-576.

Douglas, K. S., & Belfrage, H. (2014). Interrater reliability and concurrent validity of the HCR-20 Version 3. *International Journal of Forensic Mental Health, 13*, 130-139.

Douglas, K. S., Blanchard, A. J. E., Guy, L. S., Reeves, K. A., & Weir, J. (2010). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography.* Retrieved from http://kdouglas.files.wordpress.com/2007/10/hcr-20-annotated-biblio-sept-2010.pdf

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

Douglas, K. S, Hart, S. D., Webster, C.D., & Belfrage, H. (2011). *Historical Clinical Risk Management (Version 3): Professional guidelines for evaluating risk of violence [Draft 2.1]*. Vancouver, BC: Mental Health, Law, and Policy Institute, Simon Fraser University.

Douglas, K. S, Ogloff, J. R. P., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology, 67*(6), 917-930.

Douglas, K. S., Shaffer, C., Blanchard, A. J. E., Guy, L. S., Reeves, K. A., & Weir, J. (2014). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography.* Retrieved from https://kdouglas.files.wordpress.com/2014/01/hcr-20-annotated-bibliography-version-12-january-20142.pdf

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

Douglas, K. S., & Webster, C. D. (1999). The HCR-20 violence risk assessment scheme: concurrent validity in a sample of incarcerated offenders. *Criminal Justice and Behavior, 26*(1), 3-19.

Doyle, M., Archer-Power, L., Coid, J., Kallis, C., Ullrich, S., & Shaw, J. (2014). Predicting post-discharge community violence in England and Wales using the HCR-20^{V3}. *International Journal of Forensic Mental Health, 13*, 140-147.

Doyle, M., Dolan, M., & McGovern, J. (2002). The validity of North American risk assessment tools in predicting in-patient violent behavior in England. *Legal and Criminological Psychology, 7,* 141-154.

Doyle, M., Lewis, G., & Brisbane, M. (2008). Inplementing the Short-Term Assessment of Risk and Treatability (START) in a forensic mental health service. *The Psychiatrist, 32,* 406-408.

Elbogen, E. B., & Johnson, S. C. (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.

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24,827 people: Systematic review and meta-analysis. *British Medical Journal, 345,* e4692. DOI:10.1136/bmj.e4692.

Flannery, R. B., Wyshak, G., Tecce, J. J., & Flannery, G. J. (2014). Characteristics of American assaultive psychiatric patients: Review of published findings, 2000–2012. *Psychiatric Quarterly, 85,* 319-328.

Gendreau, P., Goggin, C. E., & Law, M. A. (1997). Predicting prison misconducts. *Criminal Justice and Behavior, 24,* 414-431.

Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the "unparalleled" measure of offender risk? A lesson in knowledge communication. *Criminal Justice and Behavior, 29,* 397-436.

Gendreau, P., Listwan, S. J., Kuhns, J. B., & Exum, M. L. (2014). Making prisoners accountable: Are contingency management programs the answer? *Criminal Justice and Behavior, 41,* 1079-1102.

Government of Alberta (2000). *Health information act*. Edmonton, AB: Alberta Queen's Printer.

Gray, N. S., Benson, R., Craig, R., Davies, H., Fitzgerald, S., Huckle, P., Maggs, R., Taylor, J., Trueman, M., Williams, T., & Snowden, R. J. (2011). The Short-Term Assessment of Risk and Treatability (START): A prospective study of inpatient behavior. *International Journal of Forensic Mental Health, 10,* 305-313.

Gray, N. S., Hill, C., McGleish, A., Timmons, D., MacCulloch, M. J., & Snowden, R. J. (2003). Prediction of violence and self-harm in mentally disordered offenders: A prospective study of the efficacy of HCR-20, PCL-R, and psychiatric symptomatology. *Journal of Consulting and Clinical Psychology, 71,* 443-451.

Griffith, J. J., Daffern, M., & Godber, T. (2013). Examination of the predictive validity of the Dynamic Appraisal of Situational Aggression in two mental health units. *International Journal of Mental Health Nursing, 22*(6), 485-492.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19-30.

Guy, L. S., Douglas, K. S., & Hendry, M. C. (2010). The role of psychopathic personality disorder in violence risk assessments using the HCR-20. *Journal of Personality Disorders, 24*(5), 551-580.

Hanson, R. K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence, 20,* 212-217.

Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology, 50*(3), 172-182.

Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior, 36,* 865-891.

Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences, 1,* 111-119.

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

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

Hare, R. D., & Neumann, C. S. (2008). Psychopathy as a clinical and empirical construct. *Annual Review of Clinical Psychology, 4,* 217-246.

Hare, R. D., & Neumann, C. S. (2009). Psychopathy: Assessment and forensic implications. *The Canadian Journal of Psychiatry, 54,* 791-802.

Harris, G. T., Rice, M. E., & Cormier, C. A. (2002). Prospective validation of the Violence Risk Appraisal Guide in predicting violence recidivism among forensic patients. *Law and Human Behavior, 26*(4), 377-394.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20,* 315-355.

Hart, S. D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology, 3,* 121–137.

Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *Manual for the screening version of the psychopathy checklist revised (PCL:SV).* Toronto, ON: Multi-Health Systems.

Hart, S., Sturmey, P., Logan, C., & McMurran, M. (2011). Forensic case formulation. *International Journal of Forensic Mental Health, 10*(2), 118-126.

Hill, C., Rogers, R., & Bickford M. (1996). Predicting aggressive and socially disruptive behavior in a maximum security forensic psychiatric hospital. *Journal of Forensic Sciences, 1,* 56-59.

Hogan, N., & Ennis, L. (2010). Assessing risk for forensic psychiatric inpatient violence: A meta-analysis. *Open Access Journal of Forensic Psychology, 2,* 137-147.

Humphreys, L. G., & Swets, J. A. (1991). Comparison of predictive validities measured with biserial correlations and ROCs of signal detection theory. *Journal of Applied Psychology, 76,* 316-321.

Inoue, M., Tsukano, K., Muraoka, M., Kaneko, F., & Okamura, H. (2006). Psychological impact of verbal abuse and violence by patients on nurses working in psychiatric departments. *Psychiatry and Clinical Neurosciences, 60,* 29-36.

Jacob, J. N., & Holmes, D. (2011). Working under threat: Fear and nurse-patient interactions in a forensic psychiatric setting. *Journal of Forensic Nursing, 7,* 68-77.

Johnson, M. E. (2004). Violence on inpatient psychiatric units: State of the science. *Journal of the American Psychiatric Nurses Association, 10,* 113-121.

Knoll, J. L., & Resnick, P. J. (2008). Insanity defense evaluations: Toward a model for evidence-based practice. *Brief Treatment and Crisis Intervention, 8,* 92-110.

Kötter, S., von Franqué, F., Bolzmacher, M., Eucker, S., Holzinger, B., & Müller-Isberner, R. (2014). The HCR-20V3 in Germany. *International Journal of Forensic Mental Health, 13*, 122–129.

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54*(4), 337-343.

Kroner, D. G., Mills, J. F.,& Reddon, J. R. (2005). A coffee can, factor analysis, and prediction of criminal risk: The structure of "criminality." *International Journal of Law and Psychiatry,* 360-374.

Kropp, R. P., Hart, S. D., Webster, C. D., & Eaves, D. (1995). *Manual for the Spousal Assault Risk Assessment Guide* (2nd ed.). Vancouver, BC: The British Columbia Institute Against Family Violence.

Latimer, J., & Lawrence, A. (2006). *The review board systems in Canada: Overview of results from the mentally disordered accused data collection study*. Retrieved from http://www.justice.gc.ca

Lewis, G., & Doyle, M. (2009). Risk Formulation: What are we doing and why? *International Journal of Forensic Mental Health, 8*(4), 286-292.

Lewis, K., Olver, M. E., & Wong, S. C. P. (2013). The Violence Risk Scale: Predictive validity and linking treatment changes with recidivism in a sample of high risk offenders with psychopathic traits. *Assessment, 20,* 150-164.

Lipsey, M., Landenberger, N. A., & Wilson, S. J. (2007). Effects of cognitive-behavioral programs for criminal offenders: A systematic review. *Campbell Systematic Reviews, 3*(6), 1-30.

McDermott, B. E., Edens, J. F., Quanbeck, C. D., & Busse, D. (2008). Examining the role of static and dynamic risk factors in the prediction of inpatient violence: Variable- and person-focused analyses. *Law and Human Behavior, 32,* 325-338.

McNiel, D. E., & Binder, R. L. (1994). Screening for risk of inpatient violence: Validation of an actuarial tool. *Law and Human Behavior, 18*, 579−586.

McNiel, D. E., Gregory, A. L., Lam, J. N., Binder, R. L., & Sullivan, G. R. (2003). Utility of decision support tools for assessing acute risk of violence. *Journal of Consulting and Clinical Psychology, 71*(5), 945-953.

Meehl, P. E. (1954). *Clinical vs. statistical prediction*. Minneapolis, MN: University of Minnesota Press.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194-216.

Megargee, E. I. (1976). The prediction of dangerous behavior. *Criminal Justice and Behavior, 3*, 3-22.

Miller, L. (2013). Psychological evaluations in the criminal justice system: Basic Principles and best practices. *Aggression and Violent Behavior, 18*, 83-91.

Minister of Justice. (2013). Criminal Code. Retrieved from http://laws-lois.justice.gc.ca

Morrison, E. F. (1990). Violent psychiatric inpatients in a public hospital. *Scholarly Inquiry for Nursing Practice: An International Journal, 4*, 65-82.

Motiuk, L. (1991). *Antecedents and consequences of prison adjustment: A systematic assessment and reassessment approach*. Retrieved from ProQuest Dissertations and Theses.

Neves, A. C., Goncalves, R. A., & Palma-Oliveira, J. M. (2011). Assessing risk for violent and general recidivism: A study of the HCR-20 and the PCL-R with a non-clinical sample of Portuguese offenders. *International Journal of Forensic Mental Health, 10*, 137-149.

Newby, D., & Faltin, R. (2008). The very essentials of fitness for trial assessment in Canada. *Journal of Offender Rehabilitation, 47*, 185-207.

Newton, V. M., Elbogen, E. B., Brown, C. L., Snyder, J., & Barrick, A. L. (2012). Clinical decision-making about inpatient violence risk at admission to a public-sector acute psychiatric hospital. *Journal of the American Academy of Psychiatry and the Law, 40*(2), 206-214.

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

Nicholls, T. L., Brink, J., Greaves, C., Lussier, P., & Verdun-Jones, S. (2009). Forensic psychiatric inpatients and aggression: An exploration of incidence, prevalence, severity, and interventions by gender. *International Journal of Law and Psychiatry, 32*, 23-30.

Nicholls, T. L., Ogloff, J. R. P., & Douglas, K. S. (2004). Assessing risk for violence among male and female civil psychiatric patients: the HCR-20, PCL:SV, and VSC. *Behavioral Sciences and the Law, 22*, 127-158.

Nijman, H. L. I., Muris, P., Merckelbach, H. L. G. J., Palmstierna, T., Wistedt, B., Vos, A. M., van Rixtel, A., & Allertz, W. (1999). The Staff Observation Aggression Scale-Revised (SOAS-R). *Aggressive Behavior, 25,* 197-209.

Nijman, H. L. I., Palmstierna, T., Almvik, R., & Stolker, J. J. (2005). Fifteen years of research with the Staff Observation Aggression Scale: A review. *Acta Psychiatrica Scandinavica, 111*, 12–21.

Nonstad, K., Nesset, M. B., Kroppan, E., Pedersen, T. W., Nottestad, J. A., Almvik, R., & Palmstierna, T. (2010). Predictive validity and other psychometric properties of the Short-Term Assessment of Risk and Treatability (START) in a Norwegian high secure hospital. *International Journal of Forensic Mental Health, 9*(4), 294-299.

Nuffield, J. (1982). *Parole decision-making in Canada: Research towards decision guidelines* (NN75971). Ottawa, ON: Solicitor General of Canada.

Ogloff, J., & Daffern, M. (2002). *Dynamic appraisal of situational aggression: Inpatient version*. Melbourne: Monash University and Forensicare.

Ogloff, J. R. P., & Daffern, M. (2006). The Dynamic Appraisal of Situational Aggression: An instrument to assess risk for imminent aggression in psychiatric inpatients. *Behavioral Sciences & the Law, 24,* 799–813.

Olver, M. E., Beggs Christofferson, S. M., Grace, R. C., & Wong, S. C. P. (2014). Incorporating change information into sexual offender risk assessments using the Violence Risk Scale-Sexual Offender version. *Sexual Abuse: A Journal of Research and Treatment, 26,* 472-499.

Olver, M. E., & Wong, S. C. P. (2011). A comparison of static and dynamic assessment of sexual offender risk and need in a treatment context. *Criminal Justice and Behavior, 38,* 113-126.

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale- Sexual Offender Version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*(3), 318-329.

Overall, J. E. & Gorham, D. R. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports, 10,* 799-812.

Palmstierna, T., & Wistedt, B. (1989). Risk factors for aggressive behaviour are of limited value in predicting the violent behaviour of acute involuntarily admitted patients. *Acta Psychiatrica Scandinavica, 81*(2), 152-155.

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to the addictive behaviors. *American Psychologist, 47*, 1102–1114.

Public Safety Canada. (2009). *The investigation, prosecution and correctional management of high-risk offenders: A national guide* (Report No. PS4-88/2010EPDF). Ottawa, ON: Public Safety Canada.

Quinsey, V. L. (2009). Are we there yet?: Stasis and progress in forensic psychology. *Canadian Psychology, 50*(1), 15-21.

Quinsey, V. L., & Ambtman, R. (1979). Variables affecting psychiatrists' and teachers' assessments of the dangerousness of mentally ill offenders. *Journal of Consulting and Clinical Psychology, 47*, 353-362.

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

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

Reid, W. H. (2006). Sanity evaluations and criminal responsibility. *Applied Psychology in Criminal Justice, 2*, 114-146.

Reimann, B. J. (2008). *Predicting institutional aggression and seclusion within a medium secure forensic hospital using the Psychopathy Checklist - Revised (PCL-R) and the Violence Risk Appraisal Guide (VRAG)*. Retrieved from ProQuest Dissertations and Theses.

Rice, M. E. & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior, 29*, 615-620.

Rice, M. E., Harris, G. T., & Lang, C. (2013). Validation of and revision of the VRAG and SORAG: The Violence Risk Appraisal Guide – Revised (VRAG-R). *Assessment.* Advance online publication. DOI: 10.1037/a0032878.

Rice, M. E., Harris, G. T., & Quinsey, V. L. (2002). The appraisal of violence risk. *Current Opinion in Psychiatry, 15*, 589-593.

Rippon, T. J. (2000). Aggression and violence in health care professions. *Journal of Advanced Nursing*, *31*, 452-460.

Serin, R. C., Lloyd, C. D., Helmus, L., Derkzen, D. M., & Luong, D. (2013). Does intra-individual change predict offender recidivism? Searching for the Holy Grail in assessing offender change. *Aggression and Violent Behavior, 18*(1), 32-53.

Smith, S. T., Kelley, S. E., Rulesh, A., Sörman, K., & Edens, J. F. (2014). Adapting the HCR-20^{V3} for pretrial settings. *International Journal of Forensic Mental Health, 13*, 160–171.

Starzomski, A., & Wilson, K. (2015). Development of a measure to predict short-term violence in psychiatric populations: The Imminent Risk Rating Scale. *Psychological Services, 12*(1), 1-8.

Steadman, H., & Cocozza, J. (1974). *Careers of the criminally insane*. Lexington, MA: Lexington Press.

Steadman, H., & Keveles, G. (1972). The community adjustment and criminal activity of the Baxstrom patients: 1966-1970. *The American Journal of Psychiatry, 129,* 304-310.

Steadman, H. J., Silver, E., Monahan, J., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., Grisso, T., Roth, L. H., & Banks, S. (2000). A classification tree approach to the development of actuarial violence risk assessment tools. *Law and Human Behavior*, *24*, 83-100.

Strub, D. S., Douglas, K. S., & Nicholls, T. L. (2014). The validity of version 3 of the HCR-20 violence risk assessment scheme amongst offenders and civil psychiatric patients. *International Journal of Forensic Mental Health, 13*, 148–159.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Thornton, D. (2007). *Scoring guide for Risk Matrix 2000.9/SVC.* Retrieved from http://www.bhamlive1.bham.ac.uk/Documents/college-les/psych/RM2000scoringinstructions.pdf

Vaaler, A. E., Iversen, V. C., Morken, G., Fløvig, J. C., Palmstierna, T., and Linaker, O. M. (2011). Short-term prediction of threatening and violent behaviour in an acute psychiatric intensive care unit based on patient and environment characteristics. *BMC Psychiatry, 11*. Retrieved from http://www.biomedcentral.com/1471-244X/11/44.

Viljoen, J. L., Roesch, R., Ogloff, J. R. P., & Zapf, P. A. (2003). The role of Canadian psychologists in fitness and criminal responsibility evaluations. *Canadian Psychology, 44*, 369-381.

Vitacco, M. J., Erickson, S. K., Kurus, S., & Apple, S. N. (2012a). The role of Violence Risk Assessment Guide and Historical, Clinical, Risk-20 in U.S. courts: A case law survey. *Psychology, Public Policy, and Law.* Advance online publication. DOI: 10.1037/a0025834.

Vitacco, M. J., Gonsalves, V., Tomony, J., Smith, B. E. R., & Lishner, D. A. (2012b). Can standardized measures of risk predict inpatient violence?: Combining static and dynamic variables to improve accuracy. *Criminal Justice and Behavior, 39,* 589-606.

Webster, C. D., Douglas, D., Eaves, D., & Hart, D. (1997). *HCR-20: Assessing risk for violence (Version 2).* Vancouver, BC: Mental Health, Law, & Policy Institute, Simon Fraser University.

Webster, C. D., Eaves, D., Douglas, K., & Wintrup, A. (1995*). The HCR-20 scheme: The assessment of dangerousness and risk. (Version 1).* Burnaby, BC: Simon Fraser University and Forensic Psychiatric Services Commission of British Columbia.

Webster, C. D., Martin, M.-L., Brink, J. H., Nicholls, T. L., & Desmarais, S. L. (2009). *Short-Term Assessment Risk and Treatability (START)*. Coquitlam, BC: Forensic Psychiatric Services Commission.

Wilson, C. M., Desmarais, S. L., Nicholls, T. L., & Brink, J. (2010). The role of client strengths in assessments of violence risk using the Short-Term Assessment of Risk and Treatability (START). *International Journal of Forensic Mental Health, 9*, 282–293.

Wilson, C. M., Desmarais, S. L., Nicholls, T. L., Hart, S. D., & Brink, J. (2013). Predictive validity of dynamic factors: Assessing violence risk in forensic psychiatric inpatients. *Law and Human Behavior, 37*(6), 377-388.

Wong, S. C. P., & Gordon, A. (1999-2003). *Violence Risk Scale.* Saskatoon, SK: Author.

Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the violence risk scale: A treatment-friendly violence risk tool. *Psychology, Public Policy, and Law*, *12*(3), 279-309.

Wong, S., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. (2003). *The Violence Risk Scale—Sexual Offender version (VRS–SO)*. Saskatoon, SK: Regional Psychiatric Centre and University of Saskatchewan.

Woods, P., & Almvik, R. (2002). The Brøset Violence Checklist (BVC). *Acta Psychiatrica Scandinavica*, *106* (suppl. 412), 103-105.

Woods, P., Ashley, C., Kayto, D., & Heusdens, C. (2008). Piloting violence and incident reporting measures on one acute mental health inpatient unit. *Issues in Mental Health Nursing*, *29*(5), 455-469.

Woods, P., Olver, M. E., & Mueller, M. (2015). Implementing violence and incident reporting measures on a forensic mental health unit. *Manuscript under review.*

Wormith, J. S., Olver, M. E., Stevenson, H. E., & Girard, L. (2007). The long-term prediction of offender recidivism using diagnostic, personality, and risk/need approaches to offender assessment. *Psychological Services*, *4*(4), 287-305.

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*(5), 740-767.

Appendix A

Historical Clinical Risk Management - Version 3 Items (HCR-20^{V3}; Douglas, Hart, Webster, & Belfrage, 2011).

Historical Scale
H1. Violence
H2. Other antisocial behavior
H3. Relationships
H4. Employment
H5. Substance use
H6. Major mental disorder
H7. Personality disorder
H8. Traumatic experiences
H9. Violent attitudes
H10. Treatment or supervision response
Clinical Scale
C1. Insight
C2. Violent ideation or intent
C3. Symptoms of major mental disorder
C4. Instability
C5. Treatment or supervision response
Risk management scale
R1. Professional services and plans
R2. Living situation
R3. Personal support
R4. Treatment or supervision response
R5. Stress or coping

Appendix B

Psychopathy Checklist Revised Items (PCL-R; Hare, 2003)

1. Glibness/superficial charm
2. Grandiose sense of self-worth
3. Need for stimulation
4. Pathological lying
5. Conning/manipulative
6. Lack of remorse or guilt
7. Shallow affect
8. Callous/lack of empathy
9. Parasitic lifestyle
10. Poor behavioral controls
11. Promiscuous sexual behavior
12. Early behavior problems
13. Lack of realistic goals
14. Impulsivity
15. Irresponsibility
16. Failure to accept responsibility
17. Many short-term relationships
18. Juvenile delinquency
19. Revocation of conditional release
20. Criminal versatility

Appendix C

Short-Term Assessment of Risk and Treatability Items (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009)

1. Social Skills
2. Relationships
3. Occupational
4. Recreational
5. Self-Care
6. Mental State
7. Emotional State
8. Substance Use
9. Impulse Control
10. External Triggers
11. Social Support
12. Material Resources
13. Attitudes
14. Med. Adherence
15. Rule Adherence
16. Conduct
17. Insight
18. Plans
19. Coping
20. Treatability

Appendix D

Revised Violence Risk Appraisal Guide Items (VRAG-R; Rice, Harris, & Lang, 2013)

1. Lived with both biological parents to age 16
2. Elementary school maladjustment (up to and including grade 8)
3. History of alcohol and drug problems
4. Marital status (heterosexual relationships only) at time of index offense
5. Cormier-Lang score for nonviolent convictions and charges prior to index
6. Failure on conditional release
7. Age at index offense
8. Cormier-Lang score for violent for convictions and charges prior to index
9. Number of prior admissions (of one day or more) to correctional institutions (youth detention, jail, any correctional facility) for offenses prior to the index offense
10. Conduct disorder indicators (before age 15)
11. Sex offending (considering entire history including index offense, and all offenses for which there is convincing evidence whether resulting in charges/convictions or not)
12. Antisociality

Appendix E

Violence Risk Scale Items (VRS; Wong & Gordon, 1999-2003)

Static
S1. Current Age
S2. Age at First Violent Conviction
S3. Number of Juvenile Convictions
S4. Violence throughout Lifespan
S5. Prior Release Failures/Convictions
S6. Stability of Family Upbringing
Dynamic
D1. Emotional Control
D2. Criminal Personality
D3. Criminal Attitudes
D4. Work Ethic
D5. Criminal Peers
D6. Interpersonal Aggression
D7. Emotional Control
D8. Violence During Incarceration
D9. Weapon Use
D10. Insight into Violence
D11. Mental Illness
D12. Substance Abuse
D13. Stability of Relationships
D14. Community Support
D15. Released to High Risk Situations
D16. Violence Cycle
D17. Impulsivity
D18. Cognitive Distortion
D19. Compliance with Supervision
D20. Security Level of Release Institution

Appendix F

Staff Observation Aggression Scale – Revised (SOAS-R; Nijman et al., 1999)

1. Provocation
2. Means used by the patient
3. Target of Aggression
4. Consequence(s) for victim(s)
5. Measure(s) to stop aggression

Appendix G

DEMOGRAPHIC INFORMATION

Research #: _____

FPS# (if available): _____

DOB (yy/mm/dd): _____

Date of admission:

Date of Discharge (most recent prior to 2008):

Ethnicity:

1. **Caucasian**
2. **Aboriginal**
3. **Asian**
4. **African Canadian**
5. **Other (specify) _____**

Marital Status at admission:

1. Never married
2. Divorced/ separated
3. Currently common-law/married
4. Widowed

Cognitive functioning (use any info available):_____

Highest level of education (if available): _____

CRIMINAL HISTORY/ INDEX OFFENSE

Index Offense (any violent offence ; describe most serious violent offence):

1. Non-violent
2. Violent

_____

_____

_____

Total previous charges + convictions = _____

Institutional violence history:

-**any history of** forensic psychiatric inpatient violence ? yes / no

-civil psychiatric inpatient violence ? yes / no

-prison violence ? yes / no

-other hospital violence ? yes / no

Purpose for admission/Disposition/status. Include multiple if necessary (e.g. NCR Assessment; Unfit; civil commitment):

Status:	**Dates**:
_____	_____
_____	_____
_____	_____
_____	_____

PSYCHIATRIC INFORMATION

List any Axis I diagnoses: _____
List any Axis II diagnoses: _____
Substance abuse/dependence diagnose: _____

CURRENT INPATIENT VIOLENCE INCIDENTS

Aggression is defined as, "any verbal, nonverbal, or physical behavior that was threatening (to self, others, or property), or physical behavior that actually did harm (to self, others, or property)"

Incident 1:
Apparent provocation?: _____
Means used? (e.g. verbal; a chair; hand; kick; teeth; weapon – specify) _____
_____
Target/Victim (circle): staff / copatient / object / self / other
Consequence (circle): damage to object / felt threatened / pain / visible injury / treatment required
Measures used to stop aggression? (e.g. talk, medication, held with force, seclusion):
_____

Incident 2:
Apparent provocation?: _____
Means used? (e.g. verbal; a chair; hand; kick; teeth; weapon – specify) _____
_____
Target/Victim (circle): staff / copatient / object / self / other
Consequence (circle): damage to object / felt threatened / pain / visible injury / treatment required
Measures used to stop aggression? (e.g. talk, medication, held with force, seclusion):
_____

Incident 3:
Apparent provocation?: _____
Means used? (e.g. verbal; a chair; hand; kick; teeth; weapon – specify) _____
_____
Target/Victim (circle): staff / co-patient / object / self / other
Consequence (circle): damage to object / felt threatened / pain / visible injury / treatment required
Measures used to stop aggression? (e.g. talk, medication, held with force, seclusion):
_____