

The Predictive Validity of the Two-Tiered Violence Risk
Estimates Scale (TTV) in a Long Term Follow-Up of High
Risk Federal Offenders

by

Frances P. Churcher

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements for the degree of

Master of Arts

in

Psychology

Carleton University
Ottawa, Ontario

© 2015 Frances P. Churcher

Abstract

The reduction of general and violent recidivism has long been an issue of concern within the criminal justice system in Canada. Over the past few decades, several structured risk appraisal measures have been created in order to respond to this need. The Two-Tiered Violence Risk Estimates Scale (TTV; Mills & Kroner, 2005) is a measure designed to both predict the risk of violent recidivism for an individual offender and to identify critical risk management areas. The current study examined the predictive validity of the TTV in a sample of high-risk Canadian federal offenders ($n = 120$). Scores on the TTV were compared to those of the Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993), the Statistical Information on Recidivism Scale – Revised (SIR-R1; Nuffield, 1982), and the Psychopathy Checklist – Revised (PCL-R; Hare, 2003). Approximately 53% of the sample reoffended violently, with an overall recidivism rate of 73%. While the VRAG was the strongest predictor of violent recidivism in the sample, the Actuarial Risk Estimates (ARE) scale of the TTV produced a small, significant effect. The Risk Management Indicators (RMI) produced non-significant AUC values for all recidivism outcomes. Measure comparisons using AUC values and Cox regression showed that there were no differences in predictive validity. The results of this research are discussed in the context of the validation and reliability of the TTV, and contribute to the overall risk assessment literature.

Acknowledgments

First off, I would like to thank my supervisor Dr. Jeremy Mills, and my co-supervisor, Dr. Adelle Forth, for all the hard work they have put into helping me develop this thesis into my best piece of work. Thank you to my committee members Dr. Ralph Serin and Richard Nimijean for their valued feedback on my thesis. Thank you to Dr. Tony Glover for letting me use his dataset for my thesis and for all of his valuable feedback throughout the data collection, coding, and analysis process and to Diane Nicholson for helping with the inter-rater coding. Many thanks to Carolyn Bourgeois for helping with getting me set up at CSC and for all her help with data collection. Thank you to my friends, Holly, Heidi, Vivian, and especially Meagan for the incredible support that they have given me throughout the past two years. Thank you in particular to Holly and Vivian for the amazing comments given every stage of my thesis.

Finally, thank you to my parents and my sister for all the love and support they have given me throughout the years. You've always seen the best in me and helped me through the good and the bad.

Table of Contents

Abstract.....	ii
Acknowledgments.....	iii
Table of Contents.....	iv
List of Tables.....	vi
List of Appendices.....	vii
Introduction.....	1
Literature Review.....	1
Risk Assessment by Clinicians.....	1
Risk Assessment Schemes.....	4
Risk Assessment Measures.....	12
The Statistical Information on Recidivism Scale – Revised (SIR-R1).....	12
The Psychopathy Checklist – Revised (PCL-R).....	17
Violence Risk Appraisal Guide (VRAG).....	23
Meta-Analysis of Risk Assessment.....	27
Integrated-Actuarial Approach to Risk Assessment.....	30
Two-Tiered Violence Risk Estimates Scale (TTV).....	32
Development.....	32
Structure.....	33
Validation.....	36
Research Questions and Hypotheses.....	38
Methods.....	41
Participants.....	41
Measures.....	42

Procedure.....	46
Results.....	51
Descriptive Statistics for the Sample.....	51
Reliability Analyses.....	55
Relationships Among Study Variables – Concurrent Validity of the TTV.....	58
Predictive Validity of the TTV and Study Measures.....	59
Comparative Predictive Validity of Study Measures.....	62
Item Performance.....	65
Discussion.....	70
Sample Characteristics and Reliability.....	70
Concurrent Validity.....	73
Comparative Predictive Validity.....	74
Item Performance.....	77
Conceptual Role of the TTV.....	78
Limitations.....	82
Future Research Directions and Conclusions.....	83
References.....	85
Appendices.....	103
Appendix A.....	103
Appendix B.....	109
Appendix C.....	112
Appendix D.....	113

List of Tables

Table 1: Items of the TTV	36
Table 2: Sample Characteristics.....	52
Table 3: Base Rates of Endorsement for TTV Items.....	54
Table 4: Base Rates for Recidivism Outcomes.....	55
Table 5: Inter-rater Reliability for TTV Subscales.....	56
Table 6: Inter-rater Reliability for TTV Items.....	57
Table 7: Inter-rater Reliability of Recidivism Outcomes	58
Table 8: Correlations Between TTV Subscales and Other Study Measures	59
Table 9: Predictive Accuracy of Four Risk Assessment Measures for Violent, Non-Violent, and Any Recidivism.....	61
Table 10: Z-scores for AUC Comparisons between the TTV, VRAG, SIR-R1, and PCL-R for Recidivism.....	63
Table 11: Cox Regression Models for Violent, Non-Violent, and Any Recidivism	64
Table 12: Individual Item Correlations with Violent Recidivism	66
Table 13: Individual Item Correlations with Non-Violent Recidivism	68
Table 14: Individual Item Correlations with Any Recidivism.....	69

List of Appendices

Appendix A: Coding Manual.....	103
Appendix B: Revised RMI Coding Manual.....	109
Appendix C: Items of the Violence Risk Appraisal Guide (VRAG) and the Statistical Information on Recidivism (SIR-R1).....	112
Appendix D: Correlations Between Study Measures.....	113

The Predictive Validity of the Two-Tiered Violence Risk Estimates Scale (TTV) in a Long-term Follow-up of High Risk Federal Offenders

Standardized and structured violence risk assessment emerged in the early 1990's to aid clinicians in making accurate and non-biased decisions about offender risk in order to reduce rates of general and violent reoffending. Risk assessment meets four goals: to identify an approximate level of risk for an individual (called the risk estimate), to identify contributing salient risk factors for an individual, to identify strategies that may be used to manage or minimize risk, and finally, to communicate risk information (Mills, Kroner, & Morgan, 2011). Due to an increased emphasis on risk evaluation in both forensic and clinical settings, the use of structured risk assessment measures has risen over the past decade, with over 60% of patients in general psychiatric facilities and over 80% of patients in forensic psychiatric facilities receiving structured assessments for violence (Higgins, Watts, Bindman, Slade, & Thornicroft, 2005; Khiroya, Weaver, & Maden, 2009). Due to the now widespread use of risk assessment, it is essential to evaluate all methods being used to assess these individuals in order to determine their accuracy and effectiveness. One such assessment measure, the Two-Tiered Violence Risk Estimates scale (TTV; Mills & Kroner, 2005), is currently undergoing validation to assess its predictive abilities within offender populations.

Literature Review

Risk Assessment by Clinicians

It is important to consider how clinicians, in particular those in the forensic context, use risk assessment methods within their practice. Several surveys of clinicians have been conducted in order to determine their use of structured risk assessments during

forensic decision making. In a study by Tolman and Mullendore (2003) of 164 clinicians within the United States, diplomates (clinicians who were specialized and board-certified in forensic psychology) used a greater number of assessment measures than non-diplomate clinicians. Of the non-diplomate clinicians, 9% considered themselves to be forensic psychologists and over half (53%) had completed a court-ordered evaluation for risk. In comparison, 89% of the diplomates had completed a risk evaluation for the court. Overall, diplomates tended to complete more forensic-based evaluations than other clinicians. For example, 87% of diplomates completed evaluations for testimony concerning risk level in comparison to only 45% of clinicians. Furthermore, 71% of diplomates had evaluated individuals for competency to stand trial versus 25% of clinicians. While the MMPI-2 was used at high rates by both diplomates and clinicians (74% versus 54%), the use of measures more directly connected to forensic evaluations, for example the Psychopathy Checklist – Revised (PCL-R; Hare, 2003) and the Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993; Quinsey, Harris, Rice, & Cormier, 2006), was much higher among diplomates in comparison to clinicians. For example, 67% of diplomates reported using the PCL-R in risk evaluations in comparison to just 7% of clinicians. However, the use of standardized clinical instruments in risk evaluations, such as the International Classification of Diseases (ICD-10; World Health Organization, 1992), was higher among clinicians than diplomates (13% versus 2%). When evaluating the top five choices of instrument, clinicians favoured more standard clinical assessment instruments, such as the Structured Clinical Interview for DSM-IV (SCID; First, Spitzer, Gibbon, & Williams, 1996), the ICD-10, the Minnesota Multiphasic Personality Inventory (MMPI-2; Butcher, Graham, Ben-Porath, Tellegen, &

Dahlstrom, 2003), the Rorschach inkblot test (Rorschach, 1942), and the *Diagnostic Statistical Manual* (DSM-IV; American Psychiatric Association, 1994) while diplomates employed a higher use of forensic-based instruments (those measures designed to assess individuals who are involved in the legal system), for example the PCL-R, the Historical-Clinical-Risk-20 (HCR-20 V3; Douglas, Hart, Webster, & Belfrage, 2013), and the VRAG, in their evaluations.

In another study, 152 clinicians who conducted forensic evaluations were surveyed on their risk assessment practices (Archer, Buffington-Vollum, Stredny, & Handel, 2006). Approximately 80% of the respondents ($n = 121$) identified themselves as forensic psychologists and 37 participants reported being board-certified in forensic psychology. These numbers were similar to those found by Tolman and Mullendore (2003). On average, the respondents spent approximately half their time practicing in the forensic area and spent approximately 30% of their time conducting risk evaluations. Overall, 64% of the respondents ($n = 97$) reported using risk assessment measures designed for that purpose as part of their risk evaluations while only 41% ($n = 62$) reported using sex offender risk assessment measures to assess sex offenders. Among the respondents, the most commonly used violence risk assessment measures were the PCL-R, the HCR-20, the VRAG, and the Level of Service Inventory (LSI-R; Andrews & Bonta, 1995). In addition, clinicians used a variety of structured and unstructured personality tests, neuropsychology tests, cognitive and achievement tests, clinical scales, tests of competency, and tests for malingering in their assessments.

A more recent study by Viljoen, McLachlan, and Vincent (2010) surveyed 199 psychologists about their use of violence risk assessments in evaluating juvenile and adult

offenders. The majority of clinicians surveyed ($n = 139$; approximately 70% of the sample) indicated that they used risk assessment tools within a forensic context for the purposes of determining violence or recidivism risk for offenders. Furthermore, the majority of clinicians indicated that they used psychopathy assessments within a forensic context for adult clients ($n = 110$; approximately 90.2% of the sample). However, the use of psychopathy assessments was less common in those who assessed juveniles, with only 79.2% ($n = 61$) of clinicians having used a psychopathy tool at least once to assess juveniles. The majority of clinicians reported that they believed that both structured professional judgment (SPJ) and actuarial measures could be used successfully for both adult ($n = 74$; 60.7%) and juvenile ($n = 47$; 61.0%) assessments. Clinicians were more likely to use SPJ assessments with juveniles than with adults and were less likely to use actuarial type assessments with juveniles than with adults. Overall, the above research supports the widespread use of risk assessment measures among clinicians who are specialized in forensic psychology.

Risk Assessment Schemes

Risk factors related to recidivism can be divided into two categories: static risk factors, or those risk factors that will not change, and dynamic risk factors, or those risk factors that can be changed (Andrews & Bonta, 2010). Static risk factors tend to be more related to historical aspects of an offender's behaviour or demographic information, such as the offender's age at the index offense, their past criminal behaviour, and aspects of their offending history (Andrews & Bonta, 2010). Due to the fact that these factors are historical in nature, they not often targeted by interventions in order to reduce the offender's risk of reoffending. They are better applied in a prediction model, and not in a

management model (Heilbrun, 1997). A prediction model is defined as a model of assessment that focuses on identifying factors for prediction purposes, while management models identify factors that are used to manage risk (Heilbrun, 1997). Therefore, static risk factors tend to remain constant and do not reflect potential changes in risk (Andrews & Bonta, 2010).

Dynamic risk factors, also known as criminogenic needs, are those aspects of an offender's life that can be targeted by intervention to produce change in the likelihood of recidivism (Andrews & Bonta, 2010). Examples of dynamic risk factors include the presence of antisocial associates, antisocial attitudes or behaviors, factors relating to education or employment, and factors relating to motivation and need to change. Unlike static risk factors, which are historical and cannot be modified through intervention, dynamic risk factors provide a compelling treatment target due to their changing nature, as well as an acceptable measure of risk change over time. In addition, dynamic factors are easily incorporated into both a prediction model and a risk management model (Heilbrun, 1997). Dynamic risk factors relate to the "need" part of the Risk-Need-Responsivity model, which states that it is necessary to identify the appropriate criminogenic needs to serve as targets for treatments in order to produce change in the likelihood of recidivism (Andrews, Bonta, & Hoge, 1990). Dynamic factors can be further split into stable and acute factors (Hanson & Harris, 2000; Hanson, Harris, Scott, & Helmus, 2007). Stable factors are dynamic risk factors that change slowly over an extended period of time, such as psychopathy and criminal attitudes, while acute factors are dynamic risk factors that tend to fluctuate much quicker over a shorter time period,

such as current interpersonal conflict or negative affective state (Hanson & Harris, 2000; Hanson, Harris, Scott, & Helmus, 2007).

Three types of risk assessment schemes are currently employed by clinicians in practice. The first, unstructured professional judgment (UPJ), or unstructured clinical judgment, refers to the decision making process of clinicians that is unstructured, does not use empirically-selected guiding risk factors, and that is based purely on clinical experience (Hart, 1998). Instead of being guided by risk factors, clinicians have the tendency to focus more on present clinical symptoms in order to make their assessments (Hilton & Simmons, 2001). Prior to the emergence of the first structured and empirically-assessed risk assessment tools, risk judgments were primarily made by unstructured professional judgment according to the experience of the clinician making the decision (Hanson, 2009).

There are several issues with this method of risk assessment. First, because UPJ is based solely on the professional opinion of a single clinician, there is room for bias and error to occur (Grove & Meehl, 1996). Furthermore, the subjective nature of UPJ makes it difficult to standardize risk decisions between clinicians using this method. Therefore, two different clinicians could indicate different levels of risk for a single offender as they often have different ideas as to what factors constitute a high risk rating (Rice, Harris, Lang, & Bell, 1990). In addition, it is difficult to standardize risk judgments made by a single clinician over time. UPJ has been shown to lack empirical support for its effectiveness in comparison to more structured risk assessment measures. On average, predictions made by UPJ predict recidivism at no better than chance (e.g., Ægisdóttir et

al., 2006; Bonta, 2002; Quinsey et al., 2006). UPJ continues to prevail among clinicians alongside of structured and scientifically-validated risk instruments.

Structured professional judgment (SPJ), or structured clinical judgment, describes a group of instruments that use empirically-derived risk factors to help guide a clinician's judgment of risk (Hart, 1998). SPJ differs from unstructured methods of clinical judgment due to the inclusion of risk factors that are shown to be related to recidivism in the research literature and have significant theoretical relevance. The combination of these empirical factors is then used to help guide clinicians toward a risk decision. In addition, SPJ measures also provide clinicians with guidelines for case management (Hart, 1998). Offenders are rated on a number of risk factors and placed in categories corresponding to a level of risk. The risk decision derived from SPJ is usually categorical in nature (e.g., low, moderate, or high) and is determined through the judgment of the clinicians conducting the assessment. Unlike actuarial measures, it is not related to a statistical risk probability that is obtained from the development sample for the measure. Examples of SPJ tools include the HCR-20, the Spousal Assault Risk Assessment Guide (SARA; Kropp, Hart, Webster, & Eaves, 1999), and the Sexual Violence Risk – 20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997). SPJ tools generally contain a mix of static and dynamic risk factors. The inclusion of dynamic risk factors within the assessment scheme allows clinicians to target factors that may be changed through intervention. This method of risk assessment allows for the guidance of a clinician's professional judgment through the use of theoretically-derived and empirically tested guidelines and procedures (Hart, 1998).

There are disadvantages to SPJ assessments. This method of risk assessment has been criticized as allowing clinicians to make what is still thought by some to be subjective judgments even though they are based on empirically tested risk factors (Quinsey et al., 2006). Rather than obtaining an unbiased numerical score based on the number and type of risk factors present, clinicians instead are encouraged to use their professional experience and skills to examine the risk factors present and determine an overall risk level (Skeem & Monahan, 2011). Furthermore, unlike actuarial measures, they do not provide any structure to clinicians on how to combine the risk factors in the individual measures in order to apply them in clinical practice (Skeem & Monahan, 2011).

Actuarial risk assessment is most commonly defined as risk measures that assess risk probability through algorithms composed of combinations of risk factors (Meehl, 1954). However, for the purposes of this study, actuarial risk assessment will be more broadly defined as those measures that use a combination of empirical risk factors that correspond to a probability of recidivism (Mills & Gray, 2013). The risk probability from actuarial measures is based on a representative sample of offenders (Meehl, 1954; Quinsey et al., 2006). Examples of actuarial tools include measures designed to assess general recidivism, such as the General Statistical Information on Recidivism scale (SIR; Nuffield, 1982), as well as measures designed to assess violent recidivism, such as the VRAG, the STATIC-99 (Hanson & Thornton, 1999), and the Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, Cormier, 1998). Actuarial risk assessment instruments often assess static risk factors. However, some actuarial instruments, like the LSI-R, contain items that can change over time. These tools were

created using empirically relevant risk factors from within the risk assessment literature. Items may be weighted according to a statistical formula (VRAG) or may all have equal weight (LSI-R), and adding the scores from all items results in an overall score that corresponds to an associated risk probability score. This level of risk differs from the risk level obtained using SPJ instruments because this risk probability score corresponds to a reference sample group of similar offenders that share characteristics of offenders who would obtain this score. Furthermore, this method of assessment is seen as more objective as the risk levels are determined by the offender's score based on empirical data and not by the clinician's judgment (Quinsey et al., 2006). The scores are meant to be determined solely by the score obtained as clinician override has been shown to reduce the predictive accuracy of such measures (e.g., Quinsey et al., 2006; Vrana, Sroga, & Guzzo, 2008).

Despite the criticism of the lack of individual modifications in actuarial measures, research examining professional override of the final risk judgment has shown that they are largely ineffective in risk assessment and should not be used (e.g., Abbott, 2011; Hanson & Morton-Bourgon, 2009; Quinsey et al., 2006). Professional override is the process by which a clinician takes into account factors other than those contained within an actuarial measure and uses their clinical judgment to override the risk probability given by that actuarial measure (Andrews et al., 1990). However, research has shown that professional override does not significantly increase the predictive value of actuarial measures. In a study by Wormith, Hogg, and Guzzo (2012), professional overrides using the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004) were examined in samples of 1,905 sex offenders and 24,545 non-sexual

offenders; however, the use of overrides was also high for the non-sex offenders. Over an average follow-up period of four and a half years, approximately 44% of the sex offender sample generally reoffended and approximately 12% of the same sample violently reoffended. The rate of sexual recidivism in the sex offender sample was low at approximately 4%. The rates of general reoffending were higher than those in the non-sex offender sample but there were no significant differences in the rates of violent or sexual recidivism between the two groups. Overall, the raters tended to apply overrides more to the sex offenders than the non-sex offenders for the LS/CMI for predicting all types of recidivism. The resulting decrease in predictive validity from these overrides was larger for the sex offender sample than for the non-sex offenders. The authors suggest that while professional overrides may have a place in rare circumstances, it should not be considered in actuarial risk assessment without guidelines for its use.

An important consideration when using actuarial risk assessment measures is the base rates of recidivism in the sample that is being evaluated. For some measures, such as the VRAG, the weighting of individual items was derived using base rates obtained for the development sample (Quinsey et al., 2006). Overall for violent recidivism, base rates tend to be lower than non-violent offenses, and have the tendency to produce large rates of false positives (Bonta, Harman, Hann, & Cormier, 1996). False positives are a problem with probability estimates as they might overpredict risk for an individual offender due to large margins of error (Hart & Cooke, 2013). Some researchers have posited that because actuarial tools do not account for individual differences within their schemes, there is no way for clinicians to modify the level of risk posed by an offender if there are mitigating factors at play, therefore there is significant margin of error inherent in actuarial measures

(Hart & Cooke, 2013; Hart, Michie, & Cooke, 2007). Several researchers have countered this argument by stating that those individual margins of error do not degrade the accuracy of risk decisions as was previously thought (Hanson & Howard, 2010; Harris, Lowencamp, & Hilton, 2015). Therefore, it remains to be determined if these types of limitations produce greater error than clinical overrides based on individual items as applied in the SPJ approach.

Other criticisms of actuarial tools exist. As some actuarial measures are statistically derived based on what factors appear to best predict risk, there is no theoretical basis behind items included on such measures. The lack of inclusion of dynamic risk factors has been criticized by several researchers as it does not allow for intervention planning, and therefore measures containing solely static risk factors cannot be effectively included in a risk management model (Douglas & Skeem, 2005; Heilbrun, 1997). In addition, it has been argued that actuarial measures work best when the offender in question has characteristics similar to the development sample of the measure or for more common instances of an offense (Dvorskin & Heilbrun, 2001; Sjöstedt & Grann, 2002). Therefore, risk assessments conducted with actuarial risk measures should be made alongside other measures that include factors that would allow for the risk management aspect of an offender.

Risk assessment instruments have also been developed for offense-specific risk factors. One of the most widely developed areas is in assessing recidivism risk for sex offenders. Measures that have been developed assess static risk factors for sex offending, such as the STATIC-99 and the SORAG, and dynamic risk factors, such as the ACUTE-2007 (Hanson & Harris, 2007a), the STABLE-2007 (Hanson & Harris, 2007b), and the

Sex Offenders Need Assessment Rating (SONAR; Hanson & Harris, 2001). Risk assessment measures designed to evaluate risk for intimate partner violence have also become more widespread. These measures are both static- and dynamic-based in nature, and include the SARA, the Domestic Violence Risk Assessment Guide (DVRAG; Hilton, Harris, Rice, Houghton, & Eke, 2008), and the Ontario Domestic Assault Risk Assessment (ODARA; Hilton et al., 2004). Risk assessment measures designed for assessing adolescent offenders have also become more prevalent. Examples of such measures include the Structured Assessment of Violence Risk for Youth (SAVRY; Borum, Bartel, & Forth, 2006) and the Youth Level of Service Inventory (YLSI; Hoge & Andrews, 2006).

Risk Assessment Measures

Several risk assessment measures have been created to assess both general and violent recidivism with offender populations. The following section will provide details on several commonly-used measures.

The Statistical Information on Recidivism Scale – Revised (SIR-R1)

The Correctional Service of Canada (CSC) has adopted several risk assessment instruments as part of its intake assessment scheme. One of the main risk assessment measures used by CSC as part of the Offender Intake Assessment (OIA) is the Statistical Information on Recidivism Scale (SIR; Nuffield, 1982). The SIR is primarily used by CSC during the intake process to assess the possibility of the re-integration of offenders once released out into the community after incarceration (Motiuk & Nafekh, 2001). It was developed to assist with parole decision-making within federal corrections systems

(Nuffield, 1982). The original scale devised by Nuffield was revised in 1996 (Bonta et al., 1996) in order to reflect changes in Canadian legislation to form the SIR-R1.

The SIR-R1 is an actuarial risk assessment measure containing 15 static items that focus on historical risk factors. The items are as follows: current offense, age at admission, previous incarcerations, revocation or forfeiture, act of escape, security classification, age at first adult conviction, previous convictions for assault, marital status at most recent admission, interval at risk since last offense, number of dependants at most recent admission, current total aggregate sentence, previous convictions for sex offenses, previous convictions for breaking and entering, and employment status at arrest. These items can be scored using an offender's criminal history and do not need to be obtained through an interview. Items are statistically weighted according to their relationship with recidivism using a technique known as the simple summation technique (or the Burgess method; Burgess, 1928).

Scores on the SIR-R1 range from +27 to -30 and are clustered into five risk probability categories. Of the offenders who score in the *poor risk* category (consisting of scores from -30 to -9), 1 in 3 will succeed when released into the community, while of the offenders in the *very good risk* category (consisting of scores from +6 to +27), 4 in 5 will succeed in the community based on the original development sample. The scale defines recidivism as being arrested for an indictable offense within the three year period following release.

The SIR-R1 has shown good predictive validity for both general and violent recidivism. An early validation study examined the predictive validity of the SIR-R1 in a sample of 3,267 Canadian federal offenders (Bonta et al., 1996). The composition of this

sample was similar to the one used to validate the original SIR, and was composed mainly of adult non-Aboriginal males. Overall, 48.7% of the sample reoffended. Approximately 10% of the sample had originally committed a violent offense, and 9.8% violently recidivated (base rates of violent recidivism increased to 18.6% when a more broad definition of violence was used). SIR-R1 scores obtained a small correlation with violent recidivism ($r = .15$, correlation increased to $r = .20$ when violent recidivism was more broadly defined). The scores in the study ranged from +24 to -20 and there was a moderate correlation between SIR-R1 score and general recidivism ($r = .42$). With the exception of the previous conviction for a sex offense item, all individual items on the SIR-R1 significantly predicted recidivism. Despite the fact that the previous conviction for a sex offense item was not significantly related to general or broadly defined violent recidivism, this item was able to predict violent recidivism when it was more narrowly defined. There were significant relationships between risk category and both general and violent recidivism, where individuals in the higher risk categories were more likely to generally or violently reoffend. The SIR-R1 achieved an AUC value of .74, indicating its predictive efficacy for predicting general recidivism. AUC values were smaller for the prediction of violent recidivism, with an AUC value of .64 for when violent recidivism was narrowly defined and an AUC of .65 for when violent recidivism was broadly defined. It was most effective at differentiating successes and failures for offenders in the *fair* from the *good* risk categories and in the *good* from the *very good* risk categories.

In a study by Naekh and Motiuk (2002), 6,881 Canadian non-Aboriginal male offenders were followed up for an average of three years. Their recidivism data was then analyzed with respect to their SIR-R1 scores. Overall, grouping was related to success or

failure. Approximately 44% ($n = 673$) of those individuals who were classified as the *poor* risk group reoffended, while only 6% ($n = 141$) of those rated in the *very good* risk group reoffended. Recidivism was shown to be moderately related to SIR-R1 grouping ($r = .36$). The SIR-R1 was shown to be moderately predictive of success in the community, with an AUC value of .75. A follow-up study was conducted using SIR-R1 scores and recidivism data from 11,571 male, non-Aboriginal offenders (Barnum & Gobeil, 2012). Approximately 19% of the sample recidivated, and just 6% reoffended violently over a three-year follow-up period. There were strong relationships between SIR-R1 score and both general and violent recidivism ($r = -.43$ and $r = -.36$, respectively). No significant relationship was found between SIR-R1 score and sexual reoffending. The SIR-R1 risk group membership was considered to be a good predictor of both general and violent reoffending (AUC = .71 and AUC = .68, respectively) but was not a good predictor of sexual reoffending (AUC = .57). Overall, the results obtained in this study indicate that the SIR-R1 is still an efficient measure for predicting both general and violent recidivism in populations of non-Aboriginal offenders.

Wormith and Goldstone (1984) scored 203 Canadian randomly selected male offenders using the SIR. Scores on the SIR ranged between -15 and 21. Overall, offenders who were in maximum security facilities were scored in poorer outcome categories than those who were in medium security facilities. However, they were not scored significantly different than offenders who were in minimum security facilities. Offenders who were granted parole overall had scores predicting better outcomes than those who were not granted parole. In total, 44% of the sample was granted parole, and 56% were denied parole. Overall, this sample showed results similar to those obtained in

the development sample, indicating the effectiveness of the SIR in recidivism prediction. Kroner and Loza (2001) scored the SIR for a sample of 78 male offenders along with the PCL-R, the VRAG, and the Self-Appraisal Questionnaire (SAQ; Loza, 2005) to assess their predictive validities for both general and violent recidivism. Approximately 45% of the sample reoffended and 14% violently reoffended. There was a mean SIR score in this sample of -1.6 with a range of scores from -20 to +25. SIR scores were moderately correlated with both general ($r = -.47$) and violent recidivism ($r = -.30$). In terms of prediction, the SIR obtained AUC values of .76 and .74 for general and violent recidivism, respectively.

In a more recent study by Mills and Kroner (2006a), 209 federal male offenders were scored using the PCL-R, the LSI-R, the VRAG, and the SIR in order to assess the effects of discordance, defined as the level of disagreement between multiple raters, between raters on risk assessment measures and their predictive validities. The majority of the offenders included in the sample were serving sentences for violent offenses at the time. SIR scores ranged from -21 to +21, with a mean score of .39. The SIR scores in this study correlated moderately with both general and violent recidivism ($r = .38$ and $r = .30$, respectively). For violent recidivism, the SIR obtained an AUC value of .67 in the low discordance group (where there was little disagreement in risk ratings between raters) and an AUC value of .71 in the high discordance group (where there was a lot of disagreement in risk ratings). Similarly for general recidivism, the SIR obtained AUC values of .72 and .73 for the low and high discordance groups, respectively. When using a hierarchal stepwise regression model, the SIR alone contributed to the prediction of both general and violent recidivism in the high discordance group. In the low discordance

group, the VRAG and the LSI-R added to the prediction of both general and violent recidivism in addition to the SIR. Overall, the results of this study suggest that discordance appears to have a small effect on the predictive validity of the SIR in comparison to the other measures in this study. The authors suggest that measures that contain solely historically-based and static variables may be less influenced by the difference between raters during evaluation.

There are several limitations for the use of the SIR. It is not validated for use with Aboriginal offenders, female offenders, and sex offenders, making its use limited in scope (Barnum & Gobeil, 2012; Naekh & Motiuk, 2002). Furthermore, in several studies, its predictive validity is lower for predicting violent reoffending than for general recidivism. In a study by Serin (1996) examining 79 offenders over a five year period, where 75% of the sample had committed violent offenses, there was no correlation between SIR scores and violent recidivism (base rate of violent recidivism = 10%) when looking at an average follow-up period of 30 months. However, SIR scores were significantly related to general recidivism ($r = .31$). Violent recidivism was found to only have a weak association with SIR scores once base rates and selection ratios were controlled for. However, these results are contrasted with those of Bonta and colleagues (2006) and Kroner and Loza (2001), where the SIR was found to be related to and predictive of violent recidivism. Taken together, the above research shows support for the use of the SIR-R1 for the prediction of both non-violent and violent recidivism.

The Psychopathy Checklist – Revised (PCL-R)

Although not originally designed for the purposes of risk assessment, the PCL-R (Hare, 2003) is often used as an additional indicator of traits that may cause an offender

to pose a greater risk of recidivism. The PCL is also available as a screening version (PCL:SV; Hart, Hare, & Cox, 1995) and a youth version (PCL:YV; Forth, Kosson, & Hare, 2003). The PCL-R is divided into two Factors. Factor 1 consists of core personality traits that are considered to comprise “a selfish, callous, and remorseless use of others” (Hare, 2003, p. 79) and contains two facets. Facet 1 is known as Affective and Facet 2 is known as Interpersonal. Facet 1 contains items relating to glibness/superficial charm, pathological lying, a grandiose sense of self worth, and cunning/manipulative while Facet 2 contains items relating to lack of remorse/guilt, callous/lack of empathy, emotionally shallow, and failure to accept responsibility for one’s own actions. Items in Factor 2 consists of “a chronically unstable, antisocial, and socially deviant lifestyle” (Hare, 2003, p. 79) and contains Facets 3 (Lifestyle) and 4 (Antisocial History). Facet 3 contains items demonstrating a need for stimulation, impulsivity, irresponsibility, parasitic lifestyle and a lack of realistic, long-term goals while Facet 4 demonstrates poor behavioural controls, early behavioural problems, juvenile delinquency, criminal versatility, and revocation of conditional release. There are additional items not contained within the two factors that relate to promiscuous sexual behaviour and having many short-term marital relationships. Alternate three- (Cooke & Michie, 2001) and four-factor models (Neumann, Hare, & Newman, 2007) have been proposed for the construct of psychopathy. However, when the three factors models were compared, it was found that the two-factor model endorsed by the PCL-R was more predictive of criminal offending than the other models (Walters, 2015).

The scale contains 20 items which are scored as either 0 (not present), 1 (somewhat present), or 2 (always present). The PCL-R is scored from file information,

collateral information, and from a semi-structured interview. The scores on the PCL-R are then related to the percentage of individuals within the comparison populations of male offenders, female offenders, and male forensic psychiatric patients who would receive this total score, with an associated T-score. The PCL-R has a standard error of measurement (SEM) of approximately three (Hare, 1998). A score of 30 out of 40 is considered the empirically-acceptable cut-off score for determining psychopathy (Hare, 2003). At this cut-off score, only 15.7% of offenders would score higher based on a sample population of approximately 5408 North American male offenders (Hare, 2003).

The PCL-R has been found to be related to violence and recidivism in samples of offenders (e.g., Asscher et al., 2011; Harris, Rice, & Cormier, 1991; Serin, 1996). Several commonly used risk assessment measures, such as the VRAG and the HCR-20, include the PCL-R score as part of their scoring protocol. While the occurrence of psychopathy is at approximately 1% in the general population, psychopathy can be seen in 10 to 15% of incarcerated individuals (Coid, Yang, Ullrich, Roberts, & Hare, 2009; Hare, 2003).

The PCL-R manual contains inter-rater reliability coefficients for the individual items for samples of male offenders, female offenders, and male forensic psychiatric patients. The coefficients vary according to item. For male offenders, individual item coefficients range between .41 (Failure to Accept Responsibility) to .82 (Criminal Versatility). Inter-rater coefficients are also reported for the two factors and the four facets. In the sample of male offenders, reported in the manual, the coefficient for Factor 2 (ICC = .85) was higher than that of Factor 1 (ICC = .75). Inter-rater reliability coefficients varied between the facets, with an ICC of .71 for Facet 1, ICC = .67 for Facet 2, ICC = .75 for Facet 3, and ICC = .84 for facet 4 (Hare, 2003). An additional reliability

study by Grann, Långström, Tengström, and Stålenheim (1998) examining the consistency between clinical and file evaluations of the PCL-R showed similar levels of inter-rater reliability for Factor 2 (ICC = .89) and lower levels for Factor 1 (ICC = .69). The total PCL-R scores had good levels of inter-rater reliability (ICC = .88). Further research examining the field reliability of the PCL-R has shown that total PCL-R scores are lower in clinical settings than research settings (Edens, Boccaccini, & Johnson, 2010). In addition, Factor 1 consistently demonstrated lower levels of inter-rater reliability than Factor 2 or total scores between the two settings. Therefore, it is important to consider the differences in reliability between settings when conducting research with the PCL-R.

One meta-analysis examined the relationship between antisocial behaviour and the three Hare psychopathy measures (the PCL-R, the PCL:SV, and the PCL:YV) using 95 unique studies with a total sample size of 15,826 participants (Leistico, Salekin, DeCoster, & Rogers, 2008). Moderate effect sizes were found for the association of antisocial behaviour with high scores on the total PCL measures ($d = 0.55$, taken across 94 samples), Factor 1 ($d = 0.38$, taken across 54 samples), and Factor 2 ($d = 0.60$, taken across 53 samples). Therefore, scores for the total PCL, Factor 1, and Factor 2 would be in the 71st, 65th, and 72nd percentiles, respectively, for those participants that had antisocial behaviour compared to those who did not commit antisocial acts. However, there were several moderating variables which influenced effect sizes. There was a moderating effect of gender on effect sizes where total PCL and Factor 1 effect sizes were larger in samples with higher numbers of female participants. Furthermore, the analysis showed a moderating effect of race and country, where effect sizes for total PCL and Factor 2 scores were larger in samples from outside the United States and in samples

with higher numbers of Caucasian participants. Finally, the total PCL and Factor 1 effect sizes were weakened by the independence of psychopathy assessments, where studies that had made the psychopathy ratings after the collection of outcome data had higher Factor 1 scores and lower total PCL-R scores than studies where psychopathy rating were made prior to the collection of the outcome data.

Different factors and facets may be more predictive of general and violent recidivism. In a meta-analysis by Walters (2003) using 50 effect sizes derived from 42 PCL-R studies, it was shown that Factor 2 of the PCL-R was more predictive of recidivism than Factor 1. The same pattern was noted for the prediction of institutional violence. Since the items included in Factor 2 share many similarities with items found in many other violence risk assessment instruments, it would stand to reason that Factor 2 would be the better indicator of recidivism (Kroner, Mills, & Reddon, 2005). Further research suggests that Facet 4 of the PCL-R, otherwise known as the Antisocial facet, may be the most predictive of future violence and criminal behaviour. In a study by Walters, Knight, Grann, and Dahle (2008), Facet 4 obtained significant incremental validity over the other PCL-R factors and facets across six samples of participants which contained male offenders and male and female forensic psychiatric patients. Walters and Heilbrun (2010) studied a sample of male forensic patients ($n = 216$) who were evaluated using the PCL-R. Correlational analyses revealed that Facet 4 scores were strongly related to the two item parcels that made up Facet 4: General Acting Out ($r = .85$) and Criminality ($r = .94$). All other correlations between PCL-R facets and General Acting Out and Criminality were of small to moderate strength. Similar results were seen in a second sample of 230 male offenders ($r = .83$ for General Acting Out and $r = .94$ for

Criminality). Using receiver operating characteristic (ROC) analysis, two traits that were considered to form part of General Acting Out, poor behavioural controls and early behaviour problems, were tested for their predictive abilities for violent recidivism. Poor behavioural controls approached significance for the prediction of violent recidivism (AUC = .59) and early behaviour problems were not predictive of violent recidivism in the first sample, and in the second sample, early behaviour problems were predictive of violent recidivism (AUC = .70) but poor behavioural controls were not. Therefore, certain items included in Facet 4 may be more predictive of violent recidivism than other items.

Research conducted to investigate whether psychopathic traits are more strongly related to reactive violence (violence which is defined by an impulsive reaction to an emotional event) or instrumental violence (defined as violence that is more proactive and goal-oriented; Cornell et al., 1996) has shown that individuals with a high degree of psychopathic traits tend to use instrumental violence more often than reactive violence (e.g., Cornell et al., 1996; Woodworth & Porter, 2012). A recent meta-analysis by Blais, Solodukhin, and Forth (2014) examined 53 studies (encompassing 55 unique samples) with a total sample size of 8,753 individuals to determine the relationship between the type of violence and psychopathy. Overall, psychopathy was not found to be more strongly related to instrumental violence in relation to reactive violence, as moderate effect sizes were found for both. A relationship between Factor 2 of the PCL-R and reactive violence was found, while the Interpersonal facet (Facet 2) was related to instrumental violence. Therefore, the relationship between violence and psychopathy is well-documented in the literature.

Violence Risk Appraisal Guide (VRAG)

The Violence Risk Appraisal Guide (VRAG; Harris et al., 1993; Quinsey et al., 2006) was developed in the early 1990's from a sample of 618 mentally disordered offenders detained at the Oak Ridge forensic psychiatric facility between 1965 and 1980. The offenders had committed a variety of serious offenses and the majority had committed a violent offense. These individuals were later released from the facility and therefore had the opportunity to reoffend. Recidivism data was collected from the files of the Royal Canadian Mountain Police (RCMP). Violence was coded as the occurrence of a new criminal charge for an offense against another person or if the offender was returned to a maximum security facility after the commission of a violent offense. Violent recidivism was defined as offenses that included assault (physical and sexual), armed robbery, threats, and forcible confinement but did not include other offenses such as arson or possession of a weapon. Approximately 31% of the development sample reoffended violently. Twelve of the original 50 variables were significantly related to violent recidivism and only these variables were included in the final instrument.

The VRAG consists of 12 items relating to historical risk factors: lived with both biological parents to age 16, elementary school maladjustment, history of alcohol problems, marital status at the time of offense, criminal history score for nonviolent offenses prior to the index offense, failure on prior conditional release, age at index offense, victim injury, any female victim (for index offense), meets DSM-III criteria for any personality disorder, meets DSM-III criteria for schizophrenia, and PCL-R score. If the PCL-R score is not available or if it cannot be scored, scores from the Childhood and Adolescent Taxon Scale (CATS) may be used to score this item (Quinsey et al., 2006).

The items on the VRAG are given differential weightings derived using a stepwise discriminant function analysis method originally described by Nuffield (1982). Through this method, for every plus or minus 5% difference from the base rate of reoffending (in the VRAG development sample, the base rate was 31%), a weighting of +1 or -1 is assigned to the individual item. A total risk score between -26 and +38 (with individuals around -26 being at low risk of violence and individuals around +38 being at high risk for violence) is generated for the individual. After the scores are calculated, individuals are assigned to one of nine equal-sized probability “bins” based on the total score that indicates the level of risk for the individual. Individuals in the first bin (total scores below -22) pose the least risk of failure and individuals in the ninth bin (scores above +28) pose the highest risk. Each bin covers a range of eight points and is assigned a risk probability score based on a sample of similar offenders that is used to predict the likelihood of the individual reoffending in the future at time-points of 7 years and 10 years (Quinsey et al., 2006).

The VRAG has been used to predict dichotomous violence risk over a variety of timeframes. In a study by Harris, Rice, and Camilleri (2004), the VRAG was shown to accurately predict violent recidivism in a group of institutional patients over a time span of five months after release. The VRAG has also demonstrated a capacity to predict recidivism over longer follow-up periods. For example, the VRAG was able to successfully predict reoffending in a group of sex offenders over a follow-up period of 15 years, with AUC values of .70 for both violent and non-violent offending, and an AUC value of .66 for sexual recidivism (Parent, Guay, & Knight, 2011). Furthermore, the predictive validity of the VRAG has been extensively studied within correctional

populations worldwide. Although the VRAG was originally designed for use with male forensic psychiatric patients, its use has been extended to a variety of other sample populations. Several studies have found that it has some use in predicting violence in non-forensic psychiatric patients (Doyle, Carter, Shaw, & Dolan, 2012; Harris et al., 2004). Varying levels of predictive validity have also been found in samples of female offenders (Coid et al., 2009b), individuals with intellectual disabilities (Quinsey, Book, & Skilling, 2004), and sex offenders (Harris et al., 2003; Rice & Harris, 1997).

The predictive validity of the VRAG has been evaluated in over 60 studies encompassing a variety of sample populations. The original development sample for the VRAG produced an AUC value of .76 for the prediction of violent recidivism (Quinsey et al., 2006). Several subsequent studies have shown similar evaluations of predictive validity. In a study by Kröner, Stadtland, Eidt, and Nedopil (2007), German male offenders in a forensic psychiatric facility were evaluated using the VRAG. The obtained VRAG scores were then correlated with recidivism data after their release. Approximately 38% of the sample reoffended upon release. The overall VRAG scores were significantly correlated with recidivism, and had obtained an AUC of .70, indicating that the VRAG had good predictive validity within this sample.

Through cross-validation studies, the VRAG has been found to predict risk to an equal degree in comparison to other risk assessment measures. A study by Glover, Nicholson, Hemmati, Bernfeld, and Quinsey (2002) examined the predictive efficacy of the VRAG in comparison to the PCL-R, the SIR, and the VSIR in a sample of 106 Canadian male high-risk offenders. These offenders were deemed to be highly likely to reoffend by their parole officers and were referred to the authors for the study. The index

offenses for the sample included convictions for murder, attempted murder, manslaughter, robbery, assault, dangerous driving, property-related offenses, weapons offenses, sexual offenses, and drug offenses. Violent recidivism was defined as a new arrest for robbery with a weapon, kidnapping, assault, sexual assault, manslaughter, murder, forcible confinement, and pointing a weapon, and did not include technical violations of parole, harassment, threats, or robbery without a weapon. Overall, 58.5% of the offenders reoffended, and 32.1% reoffended violently over an average period of 713 days. The mean VRAG score for the sample was 13.9. Scores for those individuals who generally and violently reoffended were higher than the mean for the sample, with mean scores of 16.4 and 17.6, respectively. The VRAG was found to be significantly moderately correlated with both general ($r = .38$) and violent recidivism ($r = .32$). All measures used in the study were found to predict risk to an equal degree. The VRAG in particular obtained a common language effect size (CLES) of .72, while the SIR and PCL-R obtained CLES of .73 and .63, respectively. Similarly, in the study by Parent et al. (2011) noted above, the VRAG was as equally effective as eight other risk assessment instruments in predicting violent and sexual recidivism. All of the instruments, save for the Rapid Risk Assessment for Sex Offender Recidivism (RRASOR; Hanson, 1997), significantly predicted violent recidivism, with AUC values ranging from .59 to .71. The AUC values for sexual recidivism ranged from .62 to .71.

The VRAG has established acceptable levels of inter-rater reliability. During its development, items that did not reach a reliability criterion of .70 for categorical variables and .80 for continuous variables were removed from the final version of the measure. The original validation study for the VRAG produced a mean correlation effect

of .90 for continuous variables and a mean kappa effect of .83 for categorical variables (Harris et al., 1993; Quinsey et al., 2006). Further studies have found an intraclass correlation of .95 for the measure (Kroner & Mills, 2001). In a study evaluating the use of the VRAG in a sex offender population, the researchers obtained an inter-rater reliability of .90 (Rice & Harris, 1997).

A revised version of the VRAG, called the VRAG-R, was recently created (Rice, Harris, & Lang, 2013). The VRAG-R was developed in a similar way to the original using a sample population of 1,261 male offenders. The individuals included in the current development population were used in the original development sample and in subsequent follow-up studies. In the revised version, several older items were replaced with new items, including replacing the total PCL-R score with the Facet 4 score of the PCL-R, called antisociality. The history of alcohol problems item was expanded to include all forms of substance abuse. Both items involving DSM diagnoses and the items any female victim and injury to victim were removed from the measure. The revised version contains 12 items: lived with both parents until age 16, substance abuse, marital status, elementary school maladjustment, nonviolent criminal history, violent criminal history, age at index offense, failure on conditional release, sex offending, number of prior admissions to correctional institutions, conduct disorder score, and antisociality. Preliminary validation studies of the VRAG-R show that it predicts violent recidivism equally to that of the previous version, with an AUC value of .76 in its validation sample (compared to an AUC value of .75 found for the VRAG for the sample population).

Meta-Analysis of Risk Assessment

Several meta-analyses have been conducted to evaluate the relative predictive validity of current risk assessment schemes in relation to both violent and general recidivism. A meta-analysis by Campbell, French, and Gendreau (2009) examined 88 studies published between 1988 and 2006 that examined the predictive validities of the HCR-20, the PCL-R, the LSI-R, the VRAG, and the SIR in predicting violent recidivism and institutional violence. Overlap between the confidence intervals of the effect sizes indicated that the effect sizes were non-significant. Effect sizes for the risk measures examined in terms of violent recidivism ranged between $Z^+ = .22$ to $.32$ (Z^+ indicates r values weighted by sample size), while effect sizes for the prediction of institutional violence ranged between $Z^+ = .14$ to $.28$ (a Z^+ value for the SIR could not be calculated for institutional violence due to the fact that only one effect size was available). The VRAG was the measure that was most predictive of violent recidivism ($Z^+ = .32$, effect size overlapped with the LSI-R and the PCL-R but not with the SIR or the HCR-20). Third generation measures, consisting of those instruments that identify dynamic risk factors, performed better than second generation measures (instruments based solely on static risk factors), and had no overlap between effect sizes. Overall, instruments that contained dynamic risk factors, such as the HCR-20, were slightly more predictive of violent reoffending than those measures that contained only static risk factors (overlap between effect sizes was minimal). However, those measures that contained static or historical risk factors were better predictors of violent behaviour within institutions.

Another meta-analysis conducted by Yang, Wong, and Coid (2010) examined the relative predictive efficacy of nine violence risk assessment instruments and their subscales: the VRAG, the HCR-20, the GSIR, the LSI/LSI-R, the PCL-R, the PCL-SV,

the Offender Group Reconviction Scale (OGRS; Copas & Marshall, 1998), the Risk Matrix 2000 for Violence scale (RM2000V; Thornton, 2000), and the Violence Risk Scale (VRS; Wong & Gordon, 2009). Using a total of 28 studies published between 1998 and 2008 that encompassed 29 studies, the overall base rate of violent recidivism was 24.9%. They found that with the exception of the PCL-R Factor 1 subscale score, all instruments predicted risk to a relatively equivalent degree, producing moderate effect sizes. The effect sizes for the HCR-20 ($d = .79$), the VRAG ($d = .68$), the PCL-R ($d = .55$), and the SIR ($d = .67$) were all significant. In addition, the HCR-20 (AUC = .71), the VRAG (AUC = .68), the PCL-R (AUC = .68), and the SIR (AUC = .71) all were moderately predictive of violent recidivism. However, only 25% of the heterogeneity in the effect sizes found for the analysis was attributed to differences between the risk instruments. The majority of variance in the effect sizes in the current analysis was explained by the methodological differences between the studies (e.g., the use of different outcome criteria, the nationalities of offenders used in the studies).

A more recent meta-analysis was conducted by Fazel, Singh, Doll, and Grann (2012). Using 73 unique samples with a total sample size of 24,847 over 13 countries, they examined the risk of violent recidivism of nine measures: the VRAG, the LSI-R, the SORAG, the STATIC-99, the PCL-R, the HCR-20, the SVR-20, the SARA, and the SAVRY. Approximately 23.7% ($n = 5879$) of the sample reoffended violently over an average follow-up period of 49.6 months. Furthermore, using odds ratios (OR), they found that measures that were specifically designed to predict violence (OR = 6.07), which included the VRAG, the HCR-20, the SAVRY, and the SARA, were more accurate in their predictions than instruments designed to predict general (OR = 2.84) or

sexual reoffending (OR = 3.88), such as the STATIC-99 or the LSI-R. Ninety-two percent of individuals who did violently reoffend had been classified as being at a moderate or a high risk to reoffend, and 41% of these individuals who were classified as either moderate or high risk reoffended. Among those who did not reoffend, 36% of them were deemed low risk. Approximately 91% of offenders who were judged as not likely to reoffend did not go on to reoffend. Therefore, even though structured risk assessments have low positive predictive values (meaning that they are effective at screening those individuals who are at the highest risk to reoffend), they also have a high rate of negative predictive values and seem to be efficient at detecting those individuals who pose the lowest risk to reoffend.

Integrated-Actuarial Approach to Risk Assessment

The use of actuarial versus clinically-based (SPJ) measures has been a widely debated topic within the literature. Some proponents of actuarial measures argue that the introduction of clinical judgment into the risk decision-making process can decrease the predictive utility of risk decisions and that clinical judgment is unnecessary and does not add value (Quinsey et al., 2006), while those who support clinical judgment methods argue that actuarial methods have limitations and lack the context to make a risk decision that is appropriate for the individual (Hart, 1998). However, substantial evidence exists for the use of both these methods within the process of risk decision-making and risk management. While clinical judgment should not replace the risk estimates obtained by actuarial methods, some measure of clinical judgment using dynamic risk factors should be included in order to assist in the process of risk management (Dvoskin & Heilbrun, 2001). Surveys of clinicians have shown that risk reports that identify both the risk

factors for an individual and the strategies that can manage and reduce the risk are highly valued, indicating that risk management methods are highly valued (Heilbrun et al., 2004; Heilbrun, O'Neill, Strohman, Bowman, & Philipson, 2000).

The integrated-actuarial approach to risk assessment integrates four qualities of risk assessment: actuarial risk estimates derived from sample populations, dynamic risk factors, recommendations for treatment and intervention, and strategies to manage risk (Mills et al., 2011). The actuarial risk estimates serve as an anchor from which risk judgments can be made with associated probabilities of reoffending over set time-points. These risk anchors have been shown to be a more accurate approach to assessing risk as they are not overestimated as is the case with more categorical judgments of risk (Mills & Kroner, 2006b; Mills et al., 2011). For example, in the study by Mills and Kroner (2006b), risk judgments given in terms of a broad categorization (e.g., low, moderate, and high) were overestimated for the offenders, where raters tended to give higher judgments of recidivism risk than the actual base rates of occurrence. This overestimation persisted even after the raters were supplied with base rate information for the offenses in question. This categorical overestimation would not exist for methods using actuarial estimates since the risk probability is obtained from the total score and is not assigned by the rater. In addition, the dynamic risk factors in this approach make it possible to develop effective intervention plans specific to the needs of the offender, and will help to manage future risk by serving as indicators of successes and failures in specific domains.

This approach helps to address several issues within the risk assessment field. First, the importance of having an anchor for risk judgments is addressed by the inclusion of an actuarial-based risk indicator section. By including this anchor as part of the

assessment, clinicians are able to give an estimate of future risk using empirically-derived data obtained from similar offenders. This risk estimate can be included in risk assessment reports in order to indicate the likelihood of risk that an offender might pose to reoffend violently. In addition, the inclusion of the dynamic risk management items without the assignment of categorical judgments of risk serve as a guide for clinicians for develop effective interventions to address the most pertinent factors in an offender's life. Rather than including subjective categories of judgment (e.g., low, medium, and high), the dynamic items present in this model allow clinicians to focus interventions by targeting only those items for which the offender needs immediate intervention or future monitoring. The use of subjective risk categories in risk communication has been heralded by some researchers as lacking a solid empirical meaning and that they are able to be interpreted in several different ways and may cause a clinician to underestimate or overestimate the risk of recidivism (Hilton, Carter, Harris, & Sharpe, 2008; Mills & Kroner, 2006b). The lack of subjective risk categories in this approach allows clinicians to focus on specific criminogenic needs rather than on assigning an inherent risk category.

Two-Tiered Violence Risk Estimates Scale (TTV)

Development

The Two-Tiered Violence Risk Estimates Scale (TTV; Mills & Kroner, 2005) was developed to address several outstanding issues within the field of risk assessment. First, it was created in order to help consolidate the process of assessment by including both a section with actuarial risk indicators and a section with dynamic risk management items. Items for the measure were chosen in a manner similar to the HCR-20, where item

selection was rationally-driven (items were chosen based on their relevance in the literature) as opposed to statistically-driven (where items were chosen due to their statistical performance). A variety of measures were examined, and items that were consistent across measures were chosen (Mills & Kroner, 2005). Approximately half of the development sample reoffended in general (48%), while approximately a third of the sample (29%) reoffended violently over an average follow-up period of three years post-release.

Structure

The TTV contains two subscales with a total of 23 items (see Table 1). The first section contains actuarial historical factors pertaining to violence risk. This section is referred to as the Actuarial Risk Estimate (ARE). It contains 10 historical risk factors that are statistically weighted according to empirical evidence: childhood antisocial behaviour, adolescent antisocial behaviour, age at first adult conviction, prior incarcerations, prior convictions for assaultive behaviour, community supervision failure, history of alcohol abuse, failure to complete high school, criminal associations, and interpersonal difficulties.

Scores for the ARE were derived using a sample of 247 male offenders, giving this section a range between 0 and 13, a mean score of 8.3, a standard deviation of 3.3, and a standard error of measurement (SEM) of 1.04 (Mills et al., 2011). Once scored, the total score for this section is associated with a risk probability at time periods of 6 months, 1 year, 2 years, and 3 years post-release. The items in this section cannot change with intervention, but can change after return to the community if the offender reoffends. The items on the ARE can be scored either through file information or through an

interview with the offender and are scored with either a 0 or a 1 with the exception of three items: age of first adult convictions, prior incarcerations, and prior convictions for assaultive behaviour. These items are scored as either a 0, a 1, or a 2. A total score of 0 on the ARE is associated with a likelihood of violent recidivism in 1 year of 1.5% and a likelihood of 2.5% at 3 years. A total score of 13 is associated with a likelihood of violent recidivism of 37% after 1 year and 55% at 3 years post-release. The ARE correlates strongly with other measures of risk, including the PCL-R (.68), the LSI-R (.79), the VRAG (.67), and the SIR (.84) (Mills et al., 2011).

Confidence intervals are calculated for the total score for the ARE by adding and subtracting 1 point (the SEM for the measure) from the total score. The 95% confidence intervals provide the range of recidivism probabilities for the offender based on their total score. An example of this range calculation is given in Mills et al. (2011). If an offender scores 6 on the ARE, he would then have a range of probable scores from 5 to 7. A score of 5 is associated with a probability of recidivism of 3.5% at six months post-release while a score of 7 would be associated with a probability of reoffending of 5.5% at the same time period. Therefore, it can be said that the probability of violent recidivism for this individual offender is between 3.5 and 5.5% within the first six months after release.

The second section of the TTV contains dynamic risk factors associated with managing risk, called the Risk Management Indicators (RMI). They were designed to be measured more than once, so as to track successes and failures on the different indicators. The items in the RMI address both the dynamic risk evaluation aspect and the risk management aspect of the integrated-actuarial model. Unlike other risk assessment instruments that utilize dynamic risk factors, the RMI does not rely on categorical risk

decisions (e.g., low, medium, and high risk). Because it does not rely on these categorical judgments, the risk judgment is based solely on the risk probabilities obtained by the ARE. The RMI contains 13 dynamic risk factors: employment, financial, substance abuse, mental health, family instability, associates, attitudes, leisure, resistance to intervention, mood, social support, environment, and stressors. Items in the RMI are scored as either a 0 (not present), a 1 (present and requires monitoring) or a 2 (present and requires intervention). If an item is present and the offender has already received some form of intervention to address this risk factor, then this item would be scored as being present and requires monitoring. If the item is present, but there have been no intervention efforts or if it is extreme and intervention has not been effective, it is scored as present and requires intervention.

Table 1

Items of the TTV

Actuarial Risk Estimates (ARE)	Risk Management Indicators (RMI)
Childhood antisocial behaviour	Employment
Adolescent antisocial behaviour	Financial
Age at first adult conviction	Substance abuse
Prior incarcerations	Mental health issues
Prior convictions for assaultive behaviour	Family instability
Community supervision failure	Associates
History of alcohol abuse	Attitudes
Failure to complete high school	Leisure
Criminal associations	Resistance to intervention
Interpersonal difficulties	Mood
	Social support
	Environment
	Stressors

Validation

The TTV has not been extensively validated. An early validation study conducted by Cheston, Mills, and Kroner (2007) examined the concurrent validity of the TTV in sample of 61 male offenders. This sample was independent from that of the development sample. The majority of this sample (86%) was convicted of violent offenses and the majority (82%) was Caucasian. Scores on the TTV were compared to those of the HCR-20, the LSI-R, and the PCL-R for the sample. In general, the mean ARE obtained ($M = 8.2$) was similar to that of the original development sample ($M = 8.30$). However, the mean RMI score in this sample ($M = 4.20$) was smaller than that of the development sample ($M = 7.90$). Because the individuals in the validation sample received intervention prior to assessment while those in the development group did not, this likely accounts for the difference in mean RMI scores between the samples. In addition the TTV correlated strongly with the other risk assessment instruments examined in the study. However there

was considerable variation in the correlations between the measures of the TTV and subscales of the other risk measures. Both the ARE and the RMI were strongly correlated with Facets 3 and 4 of the PCL-R, but failed to significantly correlate with Facets 1 and 2. The ARE of the TTV correlated strongly with the Historical items of the HCR-20 while the RMI was more strongly correlated to the Risk Management items of the HCR-20. Finally, both sections of the TTV were strongly correlated with the LSI-R.

A more recent study conducted by Mills and Gray (2013) examined the predictive validity of the TTV in a sample group of 78 Canadian male offenders who had been sentenced to incarceration in federal facilities. Participants were assessed with the both the ARE and the RMI of the TTV prior to their release. Items on the RMI were scored as either 0 (absent) or 1 (present) for the purposes of this study. Participants were also scored using the HCR-20 and the Lifestyle Criminality Screening Form (LCSF; Walters, White, & Denney, 1991) in order to get measures of concurrent validity. The HCR-20 provided a similar measure using dynamic risk factors, while the LCSF is a similar measure using static risk factors. Violent recidivism in this study was defined as a subsequent conviction for any of the following offenses: assault, uttering threats/criminal harassment, murder/manslaughter, kidnapping, contact sexual offenses, or robbery.

The ARE obtained corrected item-total correlations from .21 to .73, with a mean item-total correlation of .46. The RMI displayed similar structural reliability, with corrected item-total correlations ranging from .22 to .66, and a mean item-total correlation of .39. Both the ARE and the RMI were strongly correlated with both the total scores and the subscales scores of the HCR-20 and the LCSF.

The average follow-up period for the offenders was 4,533 days. In total, 73.1% ($n = 57$) reoffended in general and 47.4% ($n = 37$) of the sample reoffended violently. The mean score for the ARE found in the sample was 8.62 ($SD = 3.14$) and the mean score for the RMI was 8.47 ($SD = 2.75$). These mean scores were similar to those found in the Cheston et al. (2007) validation study. All of the obtained AUC values were considered large by Rice and Harris's guidelines (2005), where an AUC of .714 is considered to be large. Overall, the RMI of the TTV obtained an AUC value of .800 while the ARE obtained an AUC value of .795 for violent recidivism. In terms of general recidivism, the RMI of the TTV obtained an AUC value of .749 and the ARE obtained an AUC value of .725.

The TTV compared favourably to the existing risk assessment instruments. In comparison to the TTV, the HCR-20 and the LCSF obtained AUC values of .785 and .761, respectively, for the prediction of violent recidivism and .703 and .709, respectively, for general recidivism. The results of the above research suggest that the TTV is a comparably accurate risk assessment tool to several existing measures and should be considered for future use. Since the TTV obtained similar predictive accuracy to the HCR-20, which consists of both static and dynamic items, and the LCSF, which contains static items, this is evidence for the use of the integrated-actuarial method within violence risk assessment.

Research Questions and Hypotheses

The TTV has been developed in order to assess outstanding needs within violence risk assessment. However, due to the absence of cross-validation studies regarding the TTV, the current study was designed to address those gaps in the literature. The current

study examined several questions relating to the evaluation of the TTV in comparison to other risk assessment measures.

Question 1: What is the inter-rater reliability for the TTV?

No information exists on the inter-rater reliability of the TTV. In addition, no current information exists on the inter-rater reliability for the coding of the ARE or the RMI of the TTV. The current study aimed to generate information on the inter-rater reliability for the TTV.

Question 2: What is the long-term predictive validity of the TTV in a sample of high risk federal offenders?

Due to the recent creation of the TTV, it has not been extensively validated on populations of violent offenders. Past studies evaluating the predictive validity of the TTV (Mills & Grey, 2013) have shown that this measure shows good predictive validity in populations of offenders using shorter follow-up periods. The present study aimed to provide a comparison of the predictive validity of the TTV using a sample of high-risk violent offenders using correlational analyses and ROC curves. Based on past research, it was predicted that the TTV would be able to accurately predict the probability of violent recidivism within the sample of offenders used in the current study.

Question 3: How does the TTV compare with the VRAG, the PCL-R, and the SIR-R1 in terms of predictive validity?

No research exists regarding the comparative predictive validity of the TTV with other commonly used risk assessment measures used beyond the HCR-20 and the LCSF (Mills & Grey, 2013). The current study examined the comparative predictive validity of the TTV with regards to the PCL-R, the VRAG, and the SIR-R1 using correlational

analyses, ROC analysis, and Cox survival regression. Based on the results from past validation studies, it was predicted that the TTV will demonstrate predictive validity in keeping with the other instruments with no significant difference in measured accuracy.

Methods

Participants

The data for the current study came from archival federal male offender files used in the previous study by Glover et al. (2002). A federal sentence in Canada is defined as a sentence that is two years or greater in duration. All offenders referred for assessment prior to mid-1995 were judged to be of high risk to reoffend by their parole officers. After this point, specific referral criteria in the form of a structured assessment were put into effect (Serin, 1991). All were originally incarcerated between May 1992 and November 1996 at either Collins Bay or Bath Institutions at the time of the pre-release assessments. Participants were included in the current study if they had at least one day of opportunity to reoffend since their release to the community (either through warrant expiry, statutory release, full parole, or day parole) after their release from incarceration for the confining offense. Of the 149 offenders that were originally eligible for inclusion in the Glover et al. (2002) study, 123 met the inclusion requirement for the current study. Four cases of the original 149 files were removed due to offender death prior to release, five were lost to pardons, and 17 were deported prior to release.

As the data used in the current study was archival, informed consent was not required from the participants. All information used in the current study was obtained from CSC records and no offender contact was made. As part of the risk assessment procedures in the Glover et al. (2002) study, all offenders are asked pre-assessment if their information may be used for research purposes. All efforts have been made to ensure the confidentiality and anonymity of all offenders included within the current study. All study data was stored on a secure CSC server or in a locked cabinet in order to

ensure its confidentiality and security. Ethics approval for the current project was obtained from the ethics committees at the Department of Psychology at Carleton University and the National Headquarters at CSC.

The sample was predominantly Caucasian (78.3%), with a small number of offenders being of African Canadian, Aboriginal, or Asian origin. The mean age of the sample was 30.37 years ($SD = 7.48$) at assessment, and ranged between 19 and 48 years. The most severe index offense for approximately half of the sample (49.2%) was robbery (including armed robbery, robbery with a weapon, etc.). Approximately 23% of the sample at the index offense was incarcerated for murder (including first and second degree murder, manslaughter, and attempted murder). The most severe index offenses for the rest of the sample included property crimes (3.3%), drug crimes (0.8%), assault (including uttering threats, assault causing bodily harm, and aggravated assault, 15.0%), weapons offenses (2.5%), driving offenses (2.5%), and sexual assault (0.8%). The majority of offenders (57.5%) were released on statutory release and 25.8% of the sample was released on full parole prior to reoffending. A smaller portion of the sample was released on day parole (10.0%) and at their warrant expiry dates (6.7%).

Measures

Two-Tiered Violence Risk Estimates Scale (TTV; Mills & Kroner, 2005). The TTV consists of 10 actuarial items (the Actuarial Risk Estimate, ARE) assessing historical risk factors (e.g., age at first adult conviction, community supervision failure) and 13 risk management items (the Risk Management Indicators, RMI) that allow clinicians to determine the level of management needed based on the violence risk estimate and associated dynamic factors (e.g., employment, substance abuse). Scores on

the ARE range from 0 to 13 and are associated with a probability of violent reoffending at time-points of 6 months, 1 year, 2 years, and 3 years. Preliminary validation studies have shown support for the use of this measure in predicting violent recidivism in offender populations (Mills & Gray, 2013).

Psychopathy Checklist – Revised (PCL-R; Hare, 2003). The PCL-R contains 20 items assessing behavioural and personality constructs of psychopathy. Individual items are scored on a scale from 0 to 2, where 0 indicates that the item does not apply, 1 indicates that the item may or may not apply, and 2 indicates that the item does apply. Total scores can range from 0 to 40. Items fall into two Factors. Factor 1 represents Selfish, Callous, and Remorseless Use of Others and consists of Facet 1 (Interpersonal) and Facet 2 (Affective). Factor 2 represents Chronically Unstable and Antisocial Lifestyle and encompasses Facet 3 (Lifestyle) and Facet 4 (Antisocial). Factor 2 of the PCL-R has been shown to be more strongly predictive of recidivism (Leistico et al., 2008; Walters, 2003). The PCL-R has moderate inter-rater reliability (ICC = .71 - .85) in samples of male offenders (Hare, 2003).

The Statistical Information on Recidivism (SIR-R1; Nuffield, 1982). The SIR-R1 is an actuarial risk measure used within the CSC to assess an offender's chances of success upon parole release. It consists of 15 historical risk items, including security classification, age at admission, and current offense. Scores on the SIR-R1 range from +27 to -30, and are clustered into five risk probability categories. More negative scores indicate a decreased likelihood of success upon release. The SIR-R1 has been shown to have strong validity in the prediction of general and violent recidivism (e.g., Kroner & Loza, 2001; Mills & Kroner, 2006a)

Violence Risk Appraisal Guide (VRAG; Harris et al., 1993; Quinsey et al., 2006). The VRAG is an actuarial measure used to evaluate risk for violence. It consists of 12 historical risk items, including age at index offense, marital status, and any female victim at index offense. The scale is scored between -26 and +38 and total scores on the VRAG are divided into nine risk categories that are associated with risk probabilities for 7-year and 10-year follow-up periods based on a sample of similar offenders. The VRAG has shown good predictive validity for violent recidivism among offenders in research (e.g, Harris et al., 1993; Kroner & Mills, 2001). The VRAG has good inter-rater reliability, with a mean kappa value of .83 for categorical variables and a mean inter-class correlation of .90 for continuous variables.

Recidivism information. Recidivism information was extracted from Canadian Police Information Centre (CPIC) records via the Offender Management System (OMS) of the CSC. The index offense was defined as the offense for which the offender was incarcerated at the time of the pre-release assessments in the Glover et al. (2002) study. For the purposes of the current study, violent recidivism was defined as any new convictions post-index offense for one of the following crimes: murder/manslaughter, attempted murder, assault (common assault, causing bodily harm, aggravated assault, etc.), sexual assault, indecent exposure, weapon offenses (e.g., possession of a weapon, discharging a weapon), kidnapping, arson, and robbery (robbery, armed robbery, robbery with violence, etc.). Non-violent recidivism was defined as a new conviction for crimes against the state (treason, tax evasion, etc.), criminal negligence (impaired driving, hit and run, etc.), criminal harassment, property and fraud related crimes (theft, B&E, possession of stolen property, etc), drug possession/trafficking, or escape convictions

(e.g., failure to appear, failure to comply with recognizance). Institutional offenses (offenses committed within the institution prior to release into the community) were not included as recidivism in the current study. Furthermore, violations of parole or statutory release were not counted as recidivism.

Time-at-risk. The amount of time for which an offender had an opportunity to fail while out on release in the community was calculated in days from the day of release to the day for which they were convicted for the post-release charge, the day to which they died while on release in the community, or to the end of the follow-up period (September 30, 2014). The total follow-up period from the first offender to be released and the end of the follow-up period was 7,999 days (approximately 22 years). The total follow-up period from the last offender to be released and the end of the follow-up period was 2,294 days (approximately 6.25 years). Overall, a mean follow-up period of 6,481 days (approximately 17.75 years) was observed for the sample.

The days at-risk for any form of recidivism for the sample ranged between 4 days and 7,904 days. There was a mean of 2110.89 days ($SD = 2337.62$) at-risk in the community (approximately 5.78 years). The days at-risk for violent recidivism ranged between 79 days and 7,999 days, with a mean of 3476.89 days ($SD = 2731.52$) at-risk in the community (approximately 9.53 years). For non-violent recidivism, the time-at-risk ranged between 4 days and 7,904 days, with a mean time-at-risk of 2699.78 days (approximately 7.40 years).

At the end of the follow-up period, 72.5% of the sample ($n = 87$) had committed any type of offense leading to recidivism, 64.2% ($n = 77$) had committed a non-violent offense, and 52.5% ($n = 63$) had committed a violent offense. Two offenders had

committed offenses while on unescorted or escorted temporary absences. However, these offenses were not counted as recidivism due to the fact that the offenders did not have substantial release time in the community. The first offender to reoffend with any offense did so at 4 days post-release, with the last first reoffending at 5,623 days. There was a mean time of 916.55 days ($SD = 1017.20$) to reoffend (approximately 2.50 years).

Violent recidivism occurred between 79 and 5,020 days post-release, with a mean of 1260.35 days ($SD = 1113.47$) to reoffend (approximately 3.45 years). There was a mean to non-violently reoffend of 1192.10 days ($SD = 1352.16$; approximately 3.27 years), with the time to reoffending ranging between 4 and 7,062 days post-release.

Procedure

Data collection. Current information on offender recidivism was collected from OMS. Offender records detailing charges and convictions were accessed using a unique number for each offender via the Finger Print Service (FPS). Information on scores from risk assessment measures not scored from file information in the current study or not present in the psychological records obtained via OMS (e.g., the VRAG, PCL-R, SIR-R1) was retrieved from an archival datafile used in the Glover et al. (2002) study. A variety of records (e.g., criminal profile reports, psychological reports, progress reports) were used to code information for the TTV in order to ensure the completeness of the information used to score the measures.

Of the original 149 offender files that were referred for the Glover et al. (2002) study, 123 files met the criteria for inclusion in the present study. Three files were excluded due to a paucity of social history information available for the offenders that was needed to score the TTV. The remaining files ($n = 120$) were included in the present

study. Out of 120 files, 119 contained complete PCL-R total scores and 119 files contained total VRAG scores including the PCL-R score. One file contained a total VRAG score containing a CAT score in place of the Item 12 PCL-R total score. For the purposes of the analyses, this file was not included in future VRAG calculations. All 120 files in the study contained total SIR-R1 scores.

Data coding. Each participant file was identified with a research ID number. This number was used to locate the file and ensure the accuracy of the coding with regards to the proper file. Upon coding completion, this number was kept separate from the data. The coding manual for the current study is included in Appendix A. In the first section of the coding manual, all files were coded for the full offender name, FPS number, and the research ID number. All files were coded for date of initial release post original index offense, whether any type of recidivism occurred, the conviction date of first post-release incident, whether violent recidivism occurred, whether non-violent recidivism occurred, and the conviction date for these first incidents of recidivism. All post-release offenses were coded using the category criteria for item 20 of the PCL-R. Offenses were categorized into 19 categories: property, robbery (separated into robbery and armed robbery), drugs, assault (separated into assault, assault CBH, and assaulting a peace officer), murder, weapons, sexual (separated into sexual crimes and indecent exposure), criminal negligence, fraud, escape, kidnapping, arson, obstruction, crimes against the state, and miscellaneous. Furthermore, recidivism coding also included how many violent and non-violent offenses the offender committed post-release, and how many violent and non-violent offenses the offender committed overall throughout the course of their

record. The recidivism information was blind-coded without any knowledge of the offender's psychological risk profile.

Finally, each file was coded using the scoring manual for the TTV. The ARE was coded directly from the coding manual (Mills et al., 2011). Modifications were made to the RMI coding scheme in order to reflect that the risk factors were coded while the offenders were still incarcerated (see Appendix B). For example, for item M1, Employment, the wording for a rating of 1 was changed from "if the individual is employed on a part-time basis and is capable of full-time work," (Mills et al., 2011, p. 150) to "Individual is not employed but will be actively seeking employment upon release or has some concrete employment goals upon release." File information up to the point of release was used to code both the ARE and the RMI. Once coded, the data was entered into a datasheet on a secure CSC server.

Power analysis. Due to the generally low base rate of the occurrence of violent recidivism, it is imperative for researchers to ensure that the sample sizes used in validation studies meet acceptable numbers in order to ensure that the research has sufficient power. Of the original 149 files referred for the Glover et al. (2002) study, 123 met the current criteria for inclusion. In order to determine the appropriate sample size that would obtain sufficient power for the current study, a power analysis was conducted. Using statistics designed for ROC analysis (Hanley & McNeil, 1982), a minimum sample size of $n = 88$ would be needed in order to obtain a large AUC value of .714 (Rice & Harris, 2005) that would be statistically significant at $p < .05$ at a power level of $\beta = .20$. Therefore, the included sample size of 120 in the present study reached acceptable levels of power.

Data analysis. Data analysis was conducted using SPSS 22 and MedCalc 14 analysis software. In order to describe the characteristics of the validation sample, frequency data was calculated for a number of offender characteristics (e.g., index offenses committed, age at assessment, race, current status). The base rates of violent, non-violent, and any reoffending were calculated at the end of the follow-up period.

Receiver operating characteristic (ROC) curves were generated to examine the predictive validity of both the total and subscale scores of the TTV in the sample group for violent, non-violent, and any recidivism. The AUC (area under the curve) value generated from the ROC analysis was used to evaluate the predictive ability of the measure in terms of correctly identifying who will reoffend (Mossman, 1994; Rice & Harris, 2005). Although common language effect sizes (CLES) were used in the previous Glover et al. (2002) study, AUC values are considered to be equivalent due to their similarities in calculation (Rice & Harris, 2005). In addition, point-biserial correlations (Glass & Hopkins, 1996) were calculated in order to determine the relationship between the TTV ARE and RMI scores and dichotomous recidivism outcomes. Cox regression (Allison, 1984) was used in the current study to provide a survival analysis of dichotomous recidivism (for violent, non-violent, and any recidivism) for each measure in the sample over the total follow-up period. The use of Cox regression allows for the prediction of dichotomous recidivism by analyzing the contribution of the measures to the occurrence of recidivism using the time-at-risk as the time variable. Comparison of Cox regression models was conducted by comparing the confidence intervals of the e^b values.

Pearson correlation coefficients were calculated between the included measures and their subscales to determine the relative degree to which the measures were related to each other. Finally, the AUC values obtained for both subscales of the TTV were compared to AUC values generated for the other risk assessment measures using a comparison formula by DeLong, DeLong, and Clarke-Pearson (1988) in order to compare the predictive validities of the measures.

The TTV was coded by an additional coder for a small number of randomly-selected cases ($n = 10$) in order to gather data for the inter-rater reliability of the TTV. The secondary coder was an author on the Glover et al. (2002) study (D. Nicholson). The intraclass correlation (Shrout & Fleiss, 1978) was calculated between the coded total scores from the two raters in order to determine the level of inter-rater reliability present for scoring the measures in the current study. The intraclass correlation is a measure of consistency across a trait and was used in the current study to determine the consistency of the scoring between scorers. In addition, the individual item scores were assessed for consistency between the two raters using Cohen's kappa coefficient (Cohen, 1960). An additional independent sample of randomly-selected cases ($n = 10$) was coded by the second coder using Cohen's kappa in order to establish inter-rater reliability for the follow-up recidivism coding.

Results

Descriptive Statistics for the Sample

Descriptive information for item scores is detailed in Table 2. The mean ARE score for the sample was 8.34 ($SD = 2.58$) and the mean RMI score was 7.63 ($SD = 2.64$). Overall, ARE scores ranged between 1 and 13 and RMI scores ranged between 1 and 15. Rates of item endorsement for the TTV are presented in Table 3. The most endorsed ARE items were Failure to Complete High School (89.2%), Community Supervision Failure (80.0%), Criminal Associations (79.2%), and Adolescent Antisocial Behaviour (73.3%). Approximately 65.8% of the sample was first incarcerated between the ages of 18 and 19. Approximately 51.7% of the sample had never been previously convicted of assaultive behaviour. Of the RMI items, the Mental Health, Environment, Resistance to Intervention, and Mood items were the least endorsed in the sample. The Employment, Financial, Substance Abuse, and Leisure items were the most endorsed items.

Table 2

Sample Characteristics

Sample Characteristics	<i>M</i> (SD)	Range	<i>N</i> (%)
Age at assessment (years)	30.37 (7.48)	19 – 48	
TTV			
ARE	8.34 (2.58)	1 – 13	
RMI	7.63 (2.64)	1 – 15	
PCL-R	19.43 (6.01)	3 – 30	
Factor 1	6.59 (2.94)	0 – 13	
Factor 2	10.06 (3.54)	0 – 16	
VRAG	13.26 (8.94)	-15 – 28	
SIR-R1	-.3.36 (9.15)	-20 – 24	
Racial Composition			
Caucasian			94 (78.3)
African Canadian			16 (13.3)
Aboriginal			7 (5.8)
Asian			3 (3.1)
Type of Sentence			
Determinate			108 (90.0)
Indeterminate			12 (10.0)
Type of post-index release			
DPD			12 (10.0)
FPD			31 (25.8)
SRD			69 (57.5)
WED			8 (6.7)
Most serious index offense			
Murder/Manslaughter			27 (22.5)
Robbery			59 (49.2)
Sexual Assault			1 (0.8)
Assault			18 (15.0)
Property			4 (3.3)
Other			7 (8.9)
Current Status			
Sentence Complete			86 (71.7)
Incarcerated			10 (8.3)
Community Supervision			13 (10.8)
Deceased			11 (9.2)

Note. Actuarial Risk Estimate; RMI = Risk Management Indicators; PCL-R = Psychopathy Checklist – Revised; VRAG = Violence Risk Appraisal Guide; SIR-R1 = Statistical Information on Recidivism – Revised; DPD = day parole; FPD = full parole; SRD = statutory release; WED = warrant expiry.

PCL-R total scores for the sample ranged between 3 and 30, with a mean score for the sample of 19.43 ($SD = 6.01$). The mean Factor 1 score was 6.59 ($SD = 2.94$) and scores for the sample ranged between 0 and 13. Factor 2 scores ranged between 0 and 16, with a mean score of 10.06 ($SD = 3.54$). The sample obtained a mean total VRAG score of 13.26 ($SD = 8.94$). VRAG total scores ranged between -15 and 28. VRAG scores for the sample fell into bins 2 to 9, with the majority of scores falling into bins 6 (23.5%), 7 (37.0%), and 8 (19.3%). Finally, SIR-R1 total scores in the sample ranged between -20 and 24, with a mean score of -3.36 ($SD = 9.15$).

Table 3

Base Rates of Endorsement for TTV Items

Item	Item Endorsement (%)			
	Missing	0	1	2
A1 – Childhood Antisocial Behaviour	0	50.0	50.0	-
A2 – Adolescent Antisocial Behaviour	0	26.7	73.3	-
A3 – Age at First Incarceration	0	4.2	30.0	65.8
A4 – Prior Incarcerations	0	16.7	35.0	48.3
A5 – Prior Convictions for Assaultive Behaviour	0	51.7	35.8	12.5
A6 – Community Supervision Failure	0	20.0	80.0	-
A7 – History of Alcohol Abuse	0	30.8	69.2	-
A8 – Failure to Complete High School	0	10.8	89.2	-
A9 – Criminal Associations	0	20.0	80.0	-
A10 – Interpersonal Difficulties	2.5	62.5	35.0	-
M1 – Employment	0	14.2	80.8	5.0
M2 – Financial	0.8	15.8	24.2	59.2
M3 – Substance Abuse	0	14.2	75.0	10.8
M4 – Mental Health	0	94.2	5.8	0.0
M5 – Family Instability	0	71.7	26.7	1.7
M6 – Associates	0	19.2	40.0	40.8
M7 – Attitudes	0	27.5	49.2	23.3
M8 – Leisure	0	4.2	74.2	21.7
M9 – Resistance to Intervention	0	85.8	12.5	1.7
M10 – Mood	0	94.2	5.8	0.0
M11 – Social Support	0	82.5	10.8	6.7
M12 – Environment	0	97.5	2.5	0.0
M13 – Stressors	0	84.2	15.8	0.0

Overall, offenders committed a mean of 2.3 ($SD = 3.24$) violent offenses (with a total of 272 for the sample) and a mean of 5.9 ($SD = 9.02$) non-violent offenses (total of 711) post-release. Base rates for the follow-up offenses for the sample are presented in Table 4. The type of offenses committed most often included property offenses (41.7%),

escape offenses (40.0%), assault (37.5%), and drug offenses (26.7%). Other offenses committed post-release included fraud, robbery, sexual assault, murder, kidnapping, arson, obstruction, criminal negligence, weapons offenses, and miscellaneous offenses.

Table 4

Base Rates for Recidivism Outcomes

Type of Offense	Base Rate (%)	Total	Range
Violent			
Armed Robbery	5.0	9	0 – 3
Arson	0.8	1	0 – 1
Assault	37.5	92	0 – 5
Assault CBH	14.2	30	0 – 5
Assault Police	2.5	4	0 – 2
Indecent Exposure	0.0	0	0
Kidnapping	2.5	4	0 – 2
Murder	0.8	1	0 – 1
Robbery	20.0	83	0 – 9
Sexual Assault	2.5	4	0 – 2
Weapons	20.0	36	0 – 4
Non-Violent			
Crimes Against State	0.0	0	0
Criminal Negligence	21.7	43	0 – 4
Drug Offenses	26.7	65	0 – 6
Escape	40.0	190	0 – 17
Fraud	14.2	45	0 – 12
Miscellaneous	22.5	56	0 – 6
Obstruction of Justice	11.7	20	0 – 4
Property Crimes	41.7	292	0 – 28

Note. $n = 120$.

Reliability Analyses

Reliability analyses were conducted using a sample of 10 randomly-selected offender files that were independently rated by a second trained rater. Intraclass correlations were conducted using a two-way random model for single measures (see

Table 5). Overall, the total scores for both the ARE and the RMI had strong reliability in terms of intraclass correlations for consistency and absolute agreement ($p < .001$).

Table 5

Inter-rater Reliability for TTV Subscales

Measure	<i>k</i>	<i>r</i>	ICC		Score differences	
			Absolute Agreement	Consistency	<i>M</i> (<i>SD</i>)	Range
TTV						
ARE	10	.91***	.919***	.912***	1.0 (.74)	0 – 3
RMI	10	.92***	.908***	.920***	1.1 (.99)	0 – 3

Note. *k* = number of ratings; *r* = correlation between independent rater scores; ICC = intraclass correlation coefficient (single measures, two-way random); *M* = mean; *SD* = standard deviation; TTV = Two-tiered Violence Risk Estimates scale; ARE = Actuarial Risk Estimates; RMI = Risk Management Indicators

*** $p < .001$

The reliability of individual TTV items was examined using Cohen’s kappa (see Table 6). The majority of items obtained acceptable levels of reliability. However, several items, including items A2, A9, A10, M4, and M12, obtained levels of reliability that were below acceptable standards (kappa values less than .600).

Table 6

Inter-rater Reliability for TTV Items

Item	<i>k</i>	<i>r</i>	Kappa
ARE			
A1	9	.80**	.780*
A2	9	.16	.143
A3	10	.82**	.800**
A4	10	.90***	.756***
A5	10	1.00***	1.000***
A6	10	1.00***	1.000**
A7	9	.66	.609*
A8	10	.76**	.737*
A9	10	.38	.375
A10	9	.16	.143
RMI			
M1	10	.91***	.818**
M2	10	1.00***	1.000***
M3	10	1.00***	1.000**
M4	10	.61	.545
M5	10	1.00***	1.000**
M6	10	1.00***	1.000***
M7	10	.89***	.825***
M8	10	.75*	.630**
M9	10	.90***	.825***
M10	10	.67*	.615*
M11	10	1.00***	1.000***
M12	10	.41	.286
M13	10	.82**	.800**

Note. *k* = number of ratings; *r* = correlation between independent rater scores.

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

Reliability for recidivism scoring was conducted for the categorization of offenses into violent, non-violent, and any categories using Cohen's kappa (see Table 7). The inter-rater reliability was strong for all three recidivism outcomes ($p < .05$).

Table 7

Inter-rater Reliability of Recidivism Outcomes

Measure	<i>k</i>	<i>r</i>	Kappa
Recidivism			
Violent	10	.82**	.800**
Non-Violent	10	1.00***	1.000**
Any	10	.80**	.783*

Note. *k* = number of ratings; *r* = correlation between independent rater scores.
 * $p < .05$, ** $p < .01$, *** $p < .001$

Relationships Among Study Variables – Concurrent Validity of the TTV

Pearson correlation coefficients were generated to determine the relationship of the scores on the VRAG, SIR-R1, and the PCL-R (including Factor 1 and 2 scores) to the ARE and RMI of the TTV (see Table 8). According to Cohen's (1992) criteria, the ARE obtained large significant correlations ($p < .001$) with all study measures except for Factor 1 of the PCL-R. The RMI obtained large correlations with the VRAG and Factor 2 and moderate correlations with the SIR-R1 and the PCL-R total score (at the $p < .001$ level). The correlation between the RMI and Factor 1 was not significant. Furthermore, the ARE and the RMI were moderately correlated, $r = .37, p < .001$.

Table 8

Correlations Between TTV Subscales and Other Study Measures

	ARE	RMI
VRAG	.71***	.58***
SIR-R1	.69***	.40***
PCL-R total	.61***	.43***
PCL-R F1	.20*	.16
PCL-R F2	.73***	.52***

Note. ARE = Actuarial Risk Estimate; RMI = Risk Management Indicators; VRAG = Violence Risk Appraisal Guide; SIR-R1 = Statistical Information on Recidivism – Revised; PCL-R = Psychopathy Checklist – Revised; F1 = Factor 1; F2 = Factor 2.

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

Predictive Validity of the TTV and Study Measures

Point-biserial correlations were calculated for the ARE and RMI of the TTV in order to evaluate the relationship between the measure scores and the occurrence of recidivism (see Table 9). Overall, the ARE obtained small to moderate correlations with all three recidivism outcomes ($p < .05$). The RMI obtained a small correlation with any recidivism ($p < .05$), but failed to significantly correlate with violent or non-violent recidivism.

ROC curves were generated to evaluate the predictive accuracy of the ARE and the RMI of the TTV (see Table 9). Overall, the ARE of the TTV obtained a small effect size in the prediction of violent recidivism (AUC = .612, $p < .05$); however, it did not obtain significant AUC values for the prediction of non-violent or any recidivism. The RMI obtained a small-moderate sized AUC value for any recidivism ($p < .05$). AUC values for the RMI for the prediction of violent and non-violent recidivism were not significant.

Point-biserial correlations were generated for the remaining risk assessment measure and their subscales. The VRAG and the SIR-R1 total scores obtained moderate-large correlations for all three recidivism outcomes ($p < .05$). PCL-R total scores obtained small significant correlations with non-violent and any recidivism. Factor 1 of the PCL-R did not significantly correlate with any outcome, while Factor 2 scores obtained moderate correlations with all types of recidivism ($p < .05$).

Table 9

Predictive Accuracy of Four Risk Assessment Measures for Violent, Non-Violent, and Any Recidivism

Measure	Violent Recidivism (52.5%)		
	r_{pb}	AUC	95% CI _{BE}
TTV			
ARE	.23*	.612*	.519 – .700
RMI	.15	.583	.490 – .673
PCL-R	.17	.579	.485 – .669
Factor 1	-.06	.543	.449 – .634
Factor 2	.28**	.660**	.567 – .744
SIR-R1	.33***	.682***	.591 – .764
VRAG	.38***	.719***	.630 – .798

Measure	Non-Violent Recidivism (64.2%)			Any Recidivism (72.5%)		
	r_{pb}	AUC	95% CI _{BE}	r_{pb}	AUC	95% CI _{BE}
TTV						
ARE	.19*	.573	.480 – .663	.24**	.603	.509 – .691
RMI	.17	.605	.512 – .693	.19*	.620*	.527 – .707
PCL-R	.18**	.581	.487 – .670	.19*	.587	.493 – .676
Factor 1	<-.01	.502	.409 – .595	-.04	.525	.431 – .617
Factor 2	.29**	.642**	.549 – .728	.31***	.667**	.574 – .750
SIR-R1	.32***	.681***	.590 – .763	.36***	.712***	.622 – .791
VRAG	.27***	.618*	.524 – .705	.35***	.676**	.584 – .759

Note. TTV = Two-Tiered Violence Risk Estimates Scale; ARE = Actuarial Risk Estimate; RMI = Risk Management Indicators; PCL-R = Psychopathy Checklist – Revised; SIR-R1 = Statistical Information on Recidivism – Revised; VRAG = Violence Risk Appraisal Guide; r_{pb} = point-biserial correlation; AUC = area under the curve; 95% CI_{BE} = 95% binomial exact confidence intervals derived from DeLong et al.'s (1988) calculations using MedCalc. Base rates for recidivism are indicated in parentheses.

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

ROC analysis was conducted in order to determine the predictive accuracy of the measure scores in relation to recidivism. The VRAG obtained a large AUC value for violent recidivism, a moderate AUC for any recidivism, and a small AUC value for non-violent recidivism, all significant at the $p < .05$ level. The SIR-R1 produced significant moderate AUC values for violent and non-violent recidivism, and a large significant value for any recidivism. AUC values for the PCL-R total and Factor 1 scores were non-significant for all recidivism outcomes. AUC values for Factor 2 scores were moderate for all three types of recidivism ($p < .05$).

Comparative Predictive Validity of Study Measures

Using the formula provided by DeLong et al. (1988), a comparison of the risk measures included in the study was conducted. The results of the comparison analysis are presented in Table 10. For violent recidivism, both the ARE and the RMI differed significantly from the VRAG. For non-violent and any recidivism, the ARE differed from differed significantly from the SIR-R1. In addition, AUC values for the ARE and RMI did not significantly differ from each other for violent ($Z = 0.48, p > .05$), non-violent ($Z = 0.47, p > .05$), and any recidivism ($Z = 0.23, p > .05$). To correct for the number of multiple post-hoc comparisons, the Bonferroni correction was applied (for 12 pairwise comparisons, $p < .004$). Overall, AUC values for both the ARE and the RMI did not differ significantly from that of the VRAG, the SIR-R1, the PCL-R total, Factor 1, or Factor 2 for all recidivism outcomes after corrections were applied.

Table 10

Z-scores for AUC Comparisons between the TTV, VRAG, SIR-R1, and PCL-R for Recidivism

	Violent Recidivism		Non-Violent Recidivism		Any Recidivism	
	ARE	RMI	ARE	RMI	ARE	RMI
PCL-R	0.76	0.15	0.09	0.44	0.32	0.52
Factor 1	0.90	0.54	0.81	1.19	0.77	0.96
Factor 2	1.00	1.35	1.49	0.59	1.25	0.72
VRAG	2.21*	2.69*	0.84	0.18	1.34	0.98
SIR-R1	1.55	1.62	2.35*	1.23	2.38*	1.42

Note. ARE = Actuarial Risk Estimate; RMI = Risk Management Indicators; PCL-R = Psychopathy Checklist – Revised; VRAG = Violence Risk Appraisal Guide; SIR-R1 = Statistical Information on Recidivism – Revised.

* $p < .05$

Separate predictive models using Cox regression analysis were generated to determine the predictive validity of the study measures for all three types of recidivism over time (see Table 11). ARE scores were a significant covariate over time for the prediction of violent, non-violent, and any recidivism. For the RMI, only the model predicting any recidivism was significant ($p < .05$). Models for the VRAG, SIR-R1, and Factor 2 were significant for all three outcomes while the models for PCL-R total scores and Factor 1 were non-significant.

Comparisons between the separate regression models were conducted by examining the overlap between the confidence intervals for each model. The confidence intervals for all study measures overlapped for all three recidivism outcomes, indicating that there are no significant differences in predictive abilities between the measures.

Table 11

Cox Regression Models for Violent, Non-Violent, and Any Recidivism

Measure	Violent Recidivism				
	<i>B</i>	<i>SE</i>	Wald	e^B	95% CI
TTV ARE	.145	.053	7.441**	1.156	1.042 – 1.281
TTV RMI	.083	.047	3.074	1.087	.991 – 1.192
VRAG	.074	.019	15.344***	1.077	1.038 – 1.117
SIR-R1	.060	.016	14.396***	1.061	1.030 – 1.095
PCL-R total	.036	.022	2.826	1.037	.994 – 1.082
Factor 1	-.029	.042	.465	.972	.896 – 1.055
Factor 2	.117	.042	7.760**	1.124	1.036 – 1.221

Measure	Non-Violent Recidivism				
	<i>B</i>	<i>SE</i>	Wald	e^B	95% CI
TTV ARE	.092	.044	4.323*	1.096	1.006 – 1.194
TTV RMI	.075	.040	3.419	1.078	.996 – 1.166
VRAG	.041	.014	8.055***	1.042	1.013 – 1.071
SIR-R1	.055	.014	14.592***	1.057	1.027 – 1.087
PCL-R total	.034	.019	3.203	1.034	.997 – 1.073
Factor 1	-.002	.038	.003	.998	.927 – 1.074
Factor 2	.106	.037	8.156**	1.111	1.034 – 1.194

Measure	Any Recidivism				
	<i>B</i>	<i>SE</i>	Wald	e^B	95% CI
TTV ARE	.108	.042	6.754**	1.115	1.027 – 1.209
TTV RMI	.083	.039	4.537*	1.087	1.007 – 1.173
VRAG	.051	.014	13.190***	1.052	1.024 – 1.082
SIR-R1	.060	.014	19.750***	1.062	1.034 – 1.090
PCL-R total	.032	.017	3.392	1.032	.998 – 1.067
Factor 1	-.020	.034	.350	.980	.916 – 1.048
Factor 2	.112	.034	10.540**	1.118	1.046 – 1.196

Note. ARE = Actuarial Risk Estimate; RMI = Risk Management Indicators; VRAG = Violence Risk Appraisal Guide; SIR-R1 = Statistical Information on Recidivism – Revised; PCL-R = Psychopathy Checklist – Revised; *B* = regression coefficient; *SE* = standard error of *B*; e^B = hazard ratio; 95% CI = 95% confidence intervals for e^B values.

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

Item Performance

The item performance of the TTV in relation to violent recidivism was examined with regards to the VRAG and SIR-R1 using a method similar to Mills, Kroner, and Hemmati (2007). All items of the ARE were entered into a Cox regression equation. Overall, only Item 6 (Community Supervision Failure) contributed unique variance ($p < .05$) to the model, indicating that the majority of items on this subscale had sufficient multicollinearity between them. A Cox regression model for RMI items found that four items (Financial, Family Instability, Resistance to Intervention, and Social Support) contributed unique variance to the model at the $p < .05$ level. Cox models for the VRAG showed that three items were significant (Criminal History, Failure on Prior Conditional Release, and Victim Injury) and the SIR-R1 showed that four items contributed significantly to the variance (Current Offense, Age at Admission, Previous Convictions for Assault, and Previous Convictions for B&E).

Table 12

Individual Item Correlations with Violent Recidivism

TTV Item	r_{pb}	VRAG		SIR-R1	
		Item	r_{pb}	Item	r_{pb}
ARE		1	.160	1	-.289***
A1	.083	2	.200*	2	-.212**
A2	-.008	3	.037	3	-.181*
A3	.152	4	.117	4	-.057
A4	.204*	5	.129	5	-.208*
A5	.183*	6	.326***	6	-
A6	.275**	7	.211*	7	-.104
A7	.015	8	.302***	8	-.180*
A8	.098	9	.142	9	-.177
A9	-.017	10	.239**	10	-.035
A10	.039	11	-	11	-.102
RMI		12	.108	12	-.074
M1	.030			13	.096
M2	.257**			14	-.276**
M3	.104			15	-.047
M4	-.048				
M5	-.030				
M6	-.015				
M7	.085				
M8	.139				
M9	-.081				
M10	.023				
M11	.142				
M12	-.061				
M13	-.045				

Note. r_{pb} = point-biserial correlation; TTV = Two-Tiered Violence Risk Estimates; ARE = Actuarial Risk Estimates; RMI = Risk Management Indicators; VRAG = Violence Risk Appraisal Guide; SIR-R1 = Statistical Information on Recidivism – Revised. See Appendix C for VRAG/SIR-R1 items.

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

Point-biserial correlations were generated for the individual items of the TTV, VRAG, and SIR-R1 in relation to violent recidivism (see Table 12). The majority of items in each of the measures did not significantly correlate with violent recidivism. Of

the ARE items, only items 4, 5, and 6 were significantly correlated, and only item 2 (Financial) of the RMI significantly correlated with violent recidivism. Similar items on the VRAG and SIR-R1 also correlated with violent recidivism. Item correlations with non-violent and any recidivism were similar, with items 4 and 6 of the ARE and item 2 of the RMI producing significant correlations (see Tables 13 and 14).

Table 13

Individual Item Correlations with Non-Violent Recidivism

TTV Item	VRAG		SIR-R1		
	r_{pb}	Item	r_{pb}	Item	
ARE		1	-.085	1	-.329***
A1	.052	2	.151	2	-.129
A2	.060	3	-.005	3	-.218*
A3	.077	4	.046	4	-.047
A4	.225*	5	.303***	5	-.148
A5	.079	6	.320***	6	-
A6	.278**	7	.072	7	-.157
A7	-.010	8	.315***	8	-.109
A8	.019	9	-.125	9	-.071
A9	.104	10	.320***	10	-.047
A10	-.021	11	-	11	-.152
RMI		12	.093	12	-.291
M1	.043			13	.123
M2	.242**			14	-.190*
M3	.159			15	-.138
M4	-.111				
M5	.067				
M6	.101				
M7	.127				
M8	.056				
M9	-.051				
M10	-.111				
M11	.043				
M12	-.103				
M13	-.009				

Note. r_{pb} = point-biserial correlation; TTV = Two-Tiered Violence Risk Estimates; ARE = Actuarial Risk Estimates; RMI = Risk Management Indicators; VRAG = Violence Risk Appraisal Guide; SIR-R1 = Statistical Information on Recidivism – Revised. See Appendix C for VRAG/SIR-R1 items.

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

Table 14

Individual Item Correlations with Any Recidivism

TTV Item	r_{pb}	VRAG		SIR-R1	
		Item	r_{pb}	Item	r_{pb}
ARE		1	-.007	1	-.314***
A1	.056	2	.144	2	-.165
A2	.093	3	.031	3	-.255**
A3	.144	4	.056	4	-.007
A4	.288***	5	.314***	5	-.139
A5	.109	6	.383***	6	-
A6	.345***	7	.102	7	-.205*
A7	-.007	8	.303***	8	-.132
A8	.026	9	-.063	9	-.096
A9	.019	10	.432***	10	-.080
A10	-.006	11	-	11	-.198*
RMI		12	.137	12	-.193*
M1	.043			13	.149
M2	.277**			14	-.259**
M3	.146			15	-.160
M4	-.086				
M5	.034				
M6	.054				
M7	.147				
M8	.109				
M9	-.081				
M10	-.086				
M11	.132				
M12	-.140				
M13	.012				

Note. r_{pb} = point-biserial correlation; TTV = Two-Tiered Violence Risk Estimates; ARE = Actuarial Risk Estimates; RMI = Risk Management Indicators; VRAG = Violence Risk Appraisal Guide; SIR-R1 = Statistical Information on Recidivism – Revised. See Appendix C for VRAG/SIR-R1 items.

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

Discussion

The current study evaluated the concurrent and predictive validities of the TTV, the PCL-R, the VRAG, the SIR-R1, and their subscales in the prediction of violent, non-violent, and any recidivism using correlations, AUC values, and Cox regression. The comparison of the TTV to all three measures has never been previously examined. Furthermore, the use of Cox regression to examine the predictive validity of the TTV over time is also new in the evaluation of this instrument. In addition to validity, this study provided a preliminary assessment of the inter-rater reliability of the TTV. The ARE of the TTV was strongly correlated with all measures except Factor 1 of the PCL-R. The RMI obtained strong correlations with the VRAG and PCL-R Factor 2. Overall, the VRAG was the strongest predictor of violent recidivism, with the TTV ARE achieving a small significant predictive value. The RMI was not predictive of violent recidivism. The SIR-R1 and PCL-R Factor 2 scores were the strongest predictors of non-violent recidivism, and the ARE and RMI were non-significant. For the prediction of any recidivism, the SIR-R1 was the strongest predictor, with RMI achieving a significant moderate AUC value. The ARE was not a significant predictor. Comparative analyses using AUC values and Cox regression showed no significant differences in predictive validity between the study measures. Both the ARE and the RMI achieved a high level of reliability between trained raters.

Sample Characteristics and Reliability

The recidivism base rates in the present study substantially increased since the original follow-up period reported in Glover et al. (2002). The violence recidivism rate increased from approximately 32% to 53%, while the rate for any recidivism also saw a

similar increase (73%, up from 53% in the previous study). The increase in recidivism rates is not surprising, given the high risk nature of the sample and the longer follow-up period. The Glover et al. study reported a maximum follow-up period of approximately six years, with an average of two years, while the current study saw an average follow-up period of approximately 18 years. The Cox regression analysis in the current study showed a downward trend in reoffending for all three type of recidivism, where the majority of recidivism in the current study occurred within the first five to six years of the follow-up period, and steadily decreased after this timeframe. However, some offenders did reoffend a much longer time after release (e.g., 10 or more years post-release). Research involving follow-up periods for recidivism has shown that after 10 years post-release, recidivism rates significantly drop off (e.g., Hanson, Steffy, & Gauthier, 1993), thereby making the recidivism pattern seen in the current study consistent with those seen in the recidivism literature. The high rate of overall recidivism may be responsible for the lower correlations seen between the measures and recidivism. This is because, in terms of the statistical analysis, a base rate of approximately 73% is equivalent to a base rate of 27% (Rice & Harris, 2005). Optimal analysis assumes an ideal base rate of 50%, and as base rates deviate from this ideal (e.g., 73% in the current study), the correlations between measure scores and recidivism are attenuated, leading to the lower observed correlations in the present study (Rice & Harris, 2005).

The mean ARE score obtained in the current study was similar to those in the development sample, the Cheston et al. (2007) study, and the Mills and Gray (2013) study. The mean RMI score was also similar to that obtained in the development sample. It was smaller than that in the Mills and Gray study but larger than the mean score

obtained in Cheston et al. This difference in mean RMI scores between studies is likely due to differences in when the information for the RMI was coded. In the development sample and the current sample, offender files were coded using information prior to any interventions being undertaken, while in the Cheston et al. study, files were coded just prior to release, meaning the offenders had likely participated in institutional interventions by that point. Furthermore, in the previous studies (Cheston et al. and Mills and Gray), RMI items were coded dichotomously as either not present (coded as 0) or present (coded as 1), whereas RMI items in the development sample and the present study were coded according to the coding protocol contained in the manual (either as not present, present and requiring monitoring, or present and requiring intervention), thus accounting for differences in the mean scores.

Mean scores of the comparison measures showed little change from those in Glover et al. (2002). The mean SIR-R1 score showed the greatest change between studies, dropping from -4.39 to -3.36 in the present study. The mean score in the current study was also similar to that seen in Mills and Kroner (2006a). As the average score for federal inmates on the measure is 0, the mean SIR-R1 score still demonstrates that the current sample is more high risk than other samples (e.g., Kroner & Loza, 2001).

Average scores on the VRAG also followed a similar pattern to the SIR-R1. The average score in the current study was significantly higher than the mean score obtained in the development sample for the measure and for several subsequent validation studies (Harris et al., 1993; Kröner et al., 2007; Kroner & Mills, 2001; Parent et al., 2011). Total PCL-R and subscale scores showed consistency with several previous validation studies (e.g.,

Coid et al., 2009b; Kroner & Mills, 2001) indicating that, although high risk, the sample is not higher in psychopathy than previous samples.

The level of inter-rater reliability achieved for the ARE in the present study was similar to those of the actuarial measures in the Glover et al. (2002) study. This high level of reliability demonstrates the strength of actuarial measures, including the ARE, in terms of consistent scoring by raters, even when file information alone is used (e.g., Quinsey et al., 2006). The RMI also achieved a high rate of inter-rater reliability. However, the items of the RMI were scored once at a single time-point rather than multiple times over several time-points, and therefore may have varying levels of reliability if scored as a true dynamic measure. Furthermore, although the reliability for the total scores of these subscales was high, several of the individual items obtained low levels of reliability. This could indicate issues regarding the wording of the coding protocol for these items, as they were unable to be consistently coded between independent raters.

Concurrent Validity

Similar to Cheston et al. (2007), PCL-R Factor 1 scores did not significantly correlate with RMI scores. However, in the present study, Factor 1 scores correlated significantly with the ARE despite not being significantly correlated in Cheston et al. Correlations for both the total PCL-R and Factor 2 scores with the ARE were stronger in the present study than in Cheston et al. The VRAG and SIR-R1 have not been previously examined in relation to the TTV. Previous research involving similar measures containing static items (e.g., the LCSF used in Mills and Gray, 2013), have shown correlations of a similar magnitude to those of the present study, indicating that there is some consistency across measures using historical items in terms of their inter-

correlations. The moderate correlation seen between the ARE and RMI may be due to the comparison of historical risk factors to potentially dynamic ones, resulting in lower concurrent validity than between the ARE and other static measures. Therefore, the measures in the present study demonstrated acceptable levels of concurrent validity with each other.

Comparative Predictive Validity

Despite the TTV's strong predictive validity findings in Mills and Gray (2013), the ARE of the TTV in the current study falls below those findings seen for similar actuarial measures, such as the VRAG and SIR-R1. Although the AUC for the ARE in the prediction of violent recidivism reached statistical significance in the current study, it was a small effect. Furthermore, unlike in the Mills and Gray study, the AUC for the ARE for any recidivism did not reach statistical significance. Although the ARE was significantly correlated with all three recidivism outcomes in the present study, the values were smaller than those seen in Mills and Gray. A similar pattern was seen for the RMI; however, the RMI failed to significantly correlate with non-violent recidivism in the present study. While the base rates for recidivism and the sample composition were similar between the two studies, as both consisted of high-risk offenders, it is possible that the sample in the current study did not possess a wide enough distribution of scores in order for the ARE and the RMI to effectively discriminate between risk outcomes. Furthermore, an additional issue with the current sample is its homogeneity. Participants in the present study were not recruited to the study upon initial assessment into the federal correctional system, as was the case with the offenders in Mills and Gray, but were instead recruited into the study after a second assessment. Further research should

be conducted with offender samples using less stringent inclusion criteria (e.g., not limiting inclusion to high-risk offenders) in order get a better sense of the predictive validity of the TTV in less homogeneous samples.

Although the ARE showed a small amount of predictive accuracy for violent recidivism and produced non-significant values for non-violent and any recidivism, the ARE proved an adequate predictor in survival analyses. Overall, the ARE showed the best predictive ability for violent recidivism, where for every increase in ARE score, there was a corresponding increase of 16% for that offender to violently reoffend. For non-violent recidivism there was an increase of 10% for each unit increase in ARE score and an increase of approximately 11% for every ARE score increase for any recidivism. However, after comparisons between the Cox regression models were conducted, there were no differences between the measures in terms of prediction.

Overall, reported AUC values for the comparison measures in the current study were consistent with AUC values reported in past research. The VRAG reported the highest AUC for violent recidivism. In the original Glover et al. (2002) study, the SIR-R1 reported the largest value for violent recidivism, with the VRAG placing a close second. Both measures were strong and equal predictors of violent recidivism. Additionally, in the present study, the SIR-R1 and the VRAG were equally effective in predicting any recidivism. This finding was similar to that of the Glover et al. study. The PCL-R total and Factor 1 scores, much as in the Glover et al. study, were not significant predictors of violent recidivism or any recidivism. The PCL-R Factor 2 score remained a strong predictor of both violent and any recidivism between the two studies.

Furthermore, the correlations between measure scores and recidivism outcomes were consistent with past samples. In the Glover et al. (2002) study, the PCL-R total score was not a significant predictor for any of the three recidivism outcomes. However, this finding is inconsistent with previous research in other samples showing that PCL-R total scores are correlated with violent recidivism in offender populations (e.g., Leistico et al., 2008; Serin, 1996). PCL-R Factor 1 and Factor 2 scores obtained similar results to past research, where Factor 2 correlated strongly with violent recidivism and Factor 1 was weakly correlated with violent recidivism (e.g., Glover et al., 2002; Skeem, Monahan, & Mulvey, 2002; Skeem & Mulvey, 2001). As was the case in Glover et al., all correlations between recidivism outcomes and the VRAG and SIR-R1 remained strong.

Similar to the results seen in Mills and Gray (2013), after comparison analyses, none of the AUC values for the measures included in the current study differed significantly from those of the ARE and the RMI. In addition, ROC curves for the ARE and the RMI did not differ significantly from each other for the three recidivism outcomes. The similarities in predictive outcomes seen among the measures are consistent with prior validation studies and meta-analytic research (e.g., Campbell et al., 2009; Yang et al., 2011) showing that there are no overall differences in predictive ability between existing measures. Prior to correcting for post hoc analyses, there were a few differences between the AUC values of the measures, although the significance levels for these differences were not large enough to remain significant after corrections were applied. Prior to the corrections, differences in ROC curves were seen mainly between the ARE and VRAG for violent recidivism and the ARE and SIR-R1 for non-violent and any recidivism. These differences likely occurred due to the construct of the above

measures, where the VRAG was designed to best predict violent recidivism and the SIR-R1 was designed to best predict general forms of recidivism. Furthermore, separate Cox regression models for each of the measures showed few significant differences, as the e^b values for the measures fell within an extremely narrow range, similar to the patterns seen in the Yang et al. (2011) meta-analysis with AUC values. Therefore, the results in the current study are consistent with the research literature.

Item Performance

Further inquiry into the individual item performance of the TTV for violent recidivism showed that the majority of the ARE items did not significantly correlate with recidivism. However, a similar pattern was seen for both the VRAG and the SIR-R1. In addition, items that were similar across the three measures (e.g., failure on conditional release, which includes item 6 of the ARE, item 6 of the VRAG, and item 5 of the SIR-R1) performed similarly when correlated with the recidivism outcomes. Past research examining item performance in relation to a measure's overall predictive abilities (e.g., Coid et al., 2011; Mills et al., 2007) found that the performance of individual measure items with regards to recidivism does not seem to affect the overall predictive ability of the measure. However, it is evident, as seen in Coid et al., that the predictive abilities of the measures in the current study were dependent on a small number of items.

Several reasons for this pattern of item performance have been discussed. First, the Yang et al. (2011) meta-analysis indicated that existing risk measures may evaluate an overall pattern of antisocial behaviour. This is evidenced by the behaviour of similar items across the measures. Several of the items that were significantly correlated with recidivism (e.g., history of assault, failure on prior conditional release) have been widely

linked to antisocial and criminal behaviour in the research literature. Furthermore, Kroner et al. (2005) concluded in their study that items from existing measures are contained within several overlapping dimensions, leading to a high degree of multicollinearity between measures. As the development of the TTV was guided using existing measures such as the SIR-R1 and the VRAG, it is expected that the overlapping items within these would perform similarly with regards to recidivism prediction. This pattern of item performance indicates that the smaller AUC value seen for the ARE for violent recidivism in the present study is likely not due to the performance of individual items, but may instead be indicative of wider systematic issues with the sample (e.g., homogeneity). Further research should examine the predictive validity of the TTV with regards to individual item performance.

Conceptual Role of the TTV

The TTV fulfills the four qualities of risk assessment outlined by Mills et al. (2011). First, the ARE of the TTV was created to predict the risk for violent recidivism by using a risk estimate in the form of a probability. This fulfills the first quality of risk assessment, where a risk level for an offender is identified. In the case of the ARE, this risk level is identified using a probability derived solely from static factors. According to Bonta (2002), actuarial measures should be used in place of those requiring clinical judgment to assess offender risk due to their extensive validation. While the results from the current study found promising results for the validity of the actuarial portion of the TTV in predicting violent recidivism, further validation of the ARE will be required in order to assess its accuracy at identifying risk levels.

Second, the TTV contains both static and dynamic risk factors that can be used to evaluate which factors contribute to an offender's risk level. Static factors, such as those included in the ARE, have long shown a strong relationship with recidivism in the literature (e.g., Quinsey et al., 2006). Although the RMI does not directly contribute to the risk estimate as the ARE does, it can be used to identify crucial potentially dynamic variables that may both adversely or positively affect the risk level of an offender. Research evaluating the use of both static and dynamic risk factors in prediction has showed that dynamic factors do not significantly add incremental validity to static factors when examining predictive validity for violent reoffending (e.g., Manchak, Skeem, & Douglas, 2008), but adds incremental validity when predicting general recidivism (e.g., Brown, Amand, & Zamble, 2009). The inclusion of dynamic factors in the assessment process is essential to fulfill the Need aspect of the Risk-Need-Responsivity model (RNR; Andrews et al., 1990), where dynamic factors related to criminal behaviour, also known as criminogenic needs, are assessed in order to identify treatment targets. Furthermore, the utility of dynamic items in the risk assessment process has been well documented (e.g., Andrews et al., 1990; Douglas & Skeem, 2005), making them an invaluable inclusion. Although the current study treated the RMI items as static and therefore did not examine their potentially dynamic nature, they have potential for use in tracking change in risk over time. The potentially dynamic nature of the RMI items will have to be evaluated in future research.

The inclusion of these potentially dynamic items then leads to the third quality of risk assessment, which is the identification of strategies that can be used to manage an offender's risk. By rating these dynamic variables on a scale according to the level of

intervention required, clinicians can determine what level of management is needed for individual factors. If an offender rates highly on substance abuse and requires immediate intervention, for instance, but does not present with existing mental health conditions, clinicians can choose to focus interventions directly to manage the risk posed by the substance abuse. Furthermore, the differentiation between risk factors being present and requiring monitoring versus being present and requiring immediate intervention allows for clinicians to focus on the most salient risk factors that are directly affecting risk management at the present time in order to develop the most effective correctional programs. In the current study, RMI items were coded using a modified version of the manual coding scheme. Despite these modifications, it was difficult to determine if items were present and needing monitoring or present and needing intervention using file information. Further modifications to the coding scheme to help clarify distinctions between present items needing monitoring or intervention would be beneficial for future validation studies involving the TTV and for clinicians seeking to evaluate clients using the RMI items to make risk management decisions in the future.

Finally, the TTV helps to communicate risk by both identifying a level of risk using a statistical probability of recidivism over a particular timeframe and identifying salient risk factors that may contribute to that risk and that have the potential to be modified through intervention. Communicating risk in the manner set up using the TTV fulfills the components of risk communication identified by Heilbrun and colleagues (2000, 2004). Although the ARE does not include categorical levels of risk, as was deemed preferable by the majority of clinicians in the above studies, its ability to provide different risk estimates at multiple time-points after release from incarceration allows for

a less subjective risk assessment and avoids the overestimation issues seen with measures that only provide categorical risk judgments (Hilton et al., 2008; Mills & Kroner, 2006b). However, the framing of actuarial risk estimates (e.g., stating an offender has a 25% probability of reoffending versus a 75% probability of not reoffending) have been shown to affect release decisions (Scurich & John, 2011). Research examining various risk communication approaches in legal testimony has determined that the best method of communication depends on several factors, including the context of the testimony and the issue at hand (Scurich & John, 2012). In addition, the RMI of the TTV endorses a risk management approach, which was rated as a more favourable form of risk communication over a solely predictive approach among clinicians in Heilbrun et al. Although Heilbrun et al. did not investigate the utility of combining both predictive and management styles of risk communication, it is likely that including both approaches will improve risk communication practices among clinicians. Extending the research to risk communication using approaches that combine both risk prediction and risk management, such as the TTV, may help to further knowledge on risk communication.

While further research needs to be done as to the evaluation of its predictive validity, the TTV may prove to be a valuable measure to use in the case of both risk assessment and risk management for violent offenders. Due to the inclusion of both a prediction element and a management element, the TTV fits into both the prediction and management models outlined by Heilbrun (1997) and exemplifies the integrated-actuarial model described by Mills et al. (2011). Along with similar measures and assessment schemes, such as the STATIC-99/STABLE-2007/ACUTE-2007 model of sex offender risk assessment (Hanson & Harris, 2007a,b; Hanson & Thornton, 1999), the TTV could

be considered a singular system for both prediction and management. Rather than combining multiple measures as has been done in the past (e.g., combining the VRAG and HCR-20 to touch on both actuarial prediction and risk management aspects), or using much more comprehensive and longer measures such as the LS/CMI, the TTV theoretically provides a quicker and more convenient way for clinicians and case management workers to assess both risk prediction and management.

Limitations

Several limitations were present in the current study. The primary limitation is the use of file information to score the TTV. As seen in the literature, the use of file information can result in gaps in the dataset that are not able to be filled. In the current study, three eligible files were dropped due to insufficient information to score the TTV, and one ARE item could not be scored for approximately four offenders. The effect of limited data was further compounded by the age of the data. At the time of its collection, the CSC did not have procedures to ensure that all relevant information regarding risk was gathered. Procedures were established in the CSC later on that allowed for more thorough collection of offender information, and it is likely that later samples examining the TTV (e.g., Mills & Gray, 2013) benefitted from this. To limit the effect of missing file information in the current study, every effort was made to ensure that the information used to score the TTV was complete (e.g. collection of file information from a variety of reports).

A secondary limitation to the results in the present study is that the retrospectively-scored TTV was compared to other risk assessment measures that were prospectively scored at the time of assessment. This may have created disparities in the

results seen for the TTV, as the TTV was scored solely from file information whereas the other measures would have been scored from a combination of file information and offender interviews, leading to the other measure scores being more comprehensive due to this greater wealth of information. The results seen for the other risk measures may also be more accurate because they are free of the potential biases usually associated with the use of retrospective scores in validation studies.

A third limitation is the relatively homogenous nature of the offender sample used in the current study. The majority of the sample was Caucasian, which limits the generalizability of the results to that group. Furthermore, the sample was also homogenous in terms of risk level, as the majority of offenders were incarcerated in two institutions at the time of assessment, both of which were medium security facilities. This makes the sample different from those in the previous Mills and Gray validation study, as offenders in that sample were assessed at intake and were placed at a wider selection of institutions and security levels, therefore having a greater diversity of risk level.

Future Research Directions and Conclusions

Overall, the results of this study confirms previous research that the TTV relates conceptually well to existing risk measures, especially in terms of the static items of the ARE. The TTV had a high level of inter-reliability across raters. Although the ARE proved to be a good predictor of recidivism over time, results were mixed in terms of its discriminative predictive abilities. Further research examining the predictive validity of the TTV in more heterogeneous offender samples needs to be conducted in order to examine if sample effects are a contributing factor to the results seen in the present study. Furthermore, in order to evaluate the dynamic risk management properties of the RMI,

future research should focus on assessing these variables at multiple time-points during the follow-up period, rather than at a static pre-release period, as was the case in the present study. Results from future research involving the TTV may significantly contribute to the areas of both risk assessment and risk communication.

References

- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., ... & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341-382.
doi: 10.1177/0011000005285875
- Abbott, B. R. (2011). Throwing the baby out with the bath water: Is it time for clinical judgment to supplement actuarial risk assessment?. *Journal of the American Academy of Psychiatry and the Law, 39*, 222-230.
- Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*. Beverly Hills, California, United States: Sage.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Andrews, D. A., & Bonta, J. (1995). *The Level of Service Inventory-Revised*. Toronto, Ontario, Canada: Multi-Health Systems.
- Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Cincinnati, Ohio, United States: Anderson Publishing.
- Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*, 19-52. doi:10.1177/0093854890017001004
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *Level of Supervision/Case Management Inventory (LS/CMI): An offender assessment system*. Toronto, Ontario, Canada: Multi-Health Systems.

- 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. doi: 10.1207/s15327752jpa8701_07
- Asscher, J. J., van Vugt, E. S., Stams, G. J., Dekovic, M., Eichelsheim, I. V., & Yousfi, S. (2011). The relationship between juvenile psychopathic traits, delinquency and (violent) recidivism: A meta-analysis. *Journal of Child Psychology and Psychiatry, 52*, 1134-1143. doi:10.1111/j.1469-7610.2011.02412.x
- Barnum, G., & Gobeil, R. (2012). *Prediction of reoffense using the SIR-R1 and a proxy* (Report No. R281). Ottawa, Ontario, Canada: Correctional Service of Canada.
- Blais, J., Solodukhin, E., & Forth, A. E. (2014). A meta-analysis exploring the relationship between psychopathy and instrumental versus reactive violence. *Criminal Justice and Behavior, 41*, 797-821. doi:10.1177/0093854813519629
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk-20*. Vancouver, British Columbia, Canada: British Columbia Institute Against Family Violence.
- Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice and Behavior, 29*, 355-379. doi: 10.1177/0093854802029004002
- Bonta, J., Harman, W. G., Hann, R. G., & Cormier, R. B. (1996). Prediction of recidivism among federally sentenced offenders: A re-validation of the SIR scale. *Canadian Journal of Criminology, 38*, 61-79.
- Borum, R., Bartel, P., & Forth, A. (2006). *Structured Assessment of Violence Risk in Youth (SAVRY)*. Lutz, Florida, United States: Psychological Assessment Resources.

- Brown, S. L., Amand, M. D. S., & Zamble, E. (2009). The dynamic prediction of criminal recidivism: A three-wave prospective study. *Law and Human Behavior*, 33, 25-45. doi:10.1007/s10979-008-9139-7
- Burgess, E. W. (1928). Factors determining success or failure on parole. In A. A. Bruce (Ed.), *The workings of the indeterminate sentence law and parole in Illinois* (pp. 205-249). Springfield, Illinois: Illinois State Parole Board.
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., & Dahlstrom, W. G. (2003). *MMPI-2: Minnesota multiphasic personality inventory-2*. Minneapolis, Minnesota, United States: University of Minnesota Press.
- 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.
doi: 10.1177/00938548093333610
- Cheston, J., Mills, J. F., & Kroner, D. G. (2007, June). *The two-tiered violence risk estimates: Preliminary validity of a dynamic-actuarial approach to measuring and managing violence risk*. Poster session presented at the North American Correctional and Criminal Justice Psychology Conference, Ottawa, Canada.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46. doi:10.1177/001316446002000104
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. doi: 10.1037/0033-2909.112.1.155
- Coid, J., Yang, M., Ullrich, S., Roberts, A., & Hare, R. D. (2009a). Prevalence and correlates of psychopathic traits in the household population of Great

- Britain. *International Journal of Law and Psychiatry*, 32, 65-73. doi:
10.1016/j.ijlp.2009.01.002
- Coid, J., Yang, M., Ullrich, S., Sizmur, S., Roberts, C., Farrington, D. P., & Rogers, R. D. (2009b). Gender differences in structured risk assessment: Comparing the accuracy of five instruments. *Journal of Consulting and Clinical Psychology*, 77, 337–348. doi:10.1037/a0015155
- Coid, J. W., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Farrington, D., & Rogers, R. (2011). Most items in structured risk assessment instruments do not predict violence. *The Journal of Forensic Psychiatry & Psychology*, 22, 3-21. doi: 10.1080/14789949.2010.495990
- Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment*, 13, 171-188. doi: 10.1037/1040-3590.13.2.171
- 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*, 47C, 159–171. doi:10.1111/1467-9876.00104
- Cornell, D. G., Warren, J., Hawk, G., Stafford, E., Oram, G., & Pine, D. (1996). Psychopathy in instrumental and reactive violent offenders. *Journal of Consulting and Clinical Psychology*, 64, 783-790. doi:10.1037/0022-006X.64.4.783
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44, 837-845. doi: 10.2307/2531595

- Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20 v3: Assessing risk for violence user guide*. Vancouver, British Columbia, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law, 11*, 347-383.
doi:10.1037/1076-8971.11.3.347
- Doyle, M., Carter, S., Shaw, J., & Dolan, M. (2012). Predicting community violence from patients discharged from acute mental health units in England. *Social Psychiatry and Psychiatric Epidemiology, 47*, 627-637. doi: 10.1007/s00127-011-0366-8
- Dvoskin, J. A., & Heilbrun, K. (2001). Risk assessment and release decision-making: Toward resolving the great debate. *Journal of American Academy of Psychiatry and the Law, 29*, 6-10.
- Edens, J. F., Boccaccini, M. T., & Johnson, D. W. (2010). Inter-rater reliability of the PCL-R total and factor scores among psychopathic sex offenders: Are personality features more prone to disagreement than behavioral features?. *Behavioral Sciences & the Law, 28*, 106-119. doi: 10.1002/bsl.918
- 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

- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). *Structured Clinical Interview for DSM-IV Axis I disorders, clinical version (SCID-CV)*. Washington, D.C., United States: American Psychiatric Press, Inc.
- Forth, A. E., Kosson, D. S., & Hare, R. D. (2003). *Hare Psychopathy Checklist: Youth Version (PCL: YV)*. Toronto, Ontario, Canada: Multi-Health Systems.
- Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology (3rd Ed.)*. Needham Heights, MA, United States: Allyn & Bacon.
- Glover, A. J., Nicholson, D. E., Hemmati, T., Bernfeld, G. A., & Quinsey, V. L. (2002). A comparison of predictors of general and violent recidivism among high-risk federal offenders. *Criminal Justice and Behavior*, *29*, 235-249.
doi: 10.1177/0093854802029003001
- Grann, M., Långström, N., Tengström, A., & Stålenheim, E. G. (1998). Reliability of file-based retrospective ratings of psychopathy with the PCL-R. *Journal of Personality Assessment*, *70*, 416-426. doi: 10.1207/s15327752jpa7003_2
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293-323.
doi: 10.1037/1076-8971.2.2.293
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29-36. doi: 10.1148/radiology.143.1.7063747

- Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism* (User Report 1997-04). Ottawa, Ontario, Canada: Ministry of the Solicitor General of Canada.
- Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology/Psychologie Canadienne*, *50*, 172-182.
doi:10.1037/a0015726
- Hanson, R. K., & Harris, A. J. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior*, *27*, 6-35.
doi: 10.1177/0093854800027001002
- Hanson, R. K., & Harris, A. J. (2001). A structured approach to evaluating change among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *13*, 105-122. doi: 10.1023/A:1026600304489
- Hanson, R. K., & Harris, A. J. R. (2007a). *ACUTE-2007 scoring guide*. Ottawa, Ontario, Canada: Public Safety and Emergency Preparedness.
- Hanson, R. K., & Harris, A. J. R. (2007b). *STABLE-2007 master coding guide*. Ottawa, Ontario, Canada: Public Safety and Emergency Preparedness.
- Hanson, R. K., Harris, A. J., Scott, T. L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Research User Report No. 2007-05). Ottawa, Ontario, Canada: Public Safety Canada.
- Hanson, R. K., & Howard, P. D. (2010). Individual confidence intervals do not inform decision-makers about the accuracy of risk assessment evaluations. *Law and Human Behavior*, *34*, 275-281. doi: 10.1007/s10979-010-9227-3

- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21*, 1-21. doi: 10.1037/a0014421
- Hanson, R. K., Steffy, R. A., & Gauthier, R. (1993). Long-term recidivism of child molesters. *Journal of Consulting and Clinical Psychology, 61*, 646-652. doi: 10.1037/0022-006X.61.4.646
- Hanson, R. K., & Thornton, D. (1999). *Static 99: Improving actuarial risk assessments for sex offenders (Vol. 2)*. Ottawa, Ontario, Canada: Solicitor General Canada.
- Hare, R. D. (1998). The Hare PCL-R: Some issues concerning its use and misuse. *Legal and Criminological Psychology, 3*, 99-119. doi: 10.1111/j.2044-8333.1998.tb00353.x
- Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised: PCL-R manual (2nd Ed.)*. Toronto, Ontario, Canada: Multi-Health Systems.
- Harris, G. T., Lowenkamp, C. T., & Hilton, N. Z. (2015). Evidence for risk estimate precision: Implications for individual risk communication. *Behavioral Sciences & the Law, 33*, 111-127. doi: 10.1002/bsl.2158
- Harris, G. T., Rice, M. E., & Cormier, C. A. (1991). Psychopathy and violent recidivism. *Law and Human Behavior, 15*, 625-637. doi: 10.1007/BF01065856
- Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*, 413-425. doi: 10.1037/1040-3590.15.3.413

- Harris, G. T., Rice, M. E., & Camilleri, J. A. (2004). Applying a forensic actuarial assessment (the Violence Risk Appraisal Guide) to nonforensic patients. *Journal of Interpersonal Violence, 19*, 1063-1074. doi: 10.1177/0886260504268004
- 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*, 315-335.
doi: 10.1177/0093854893020004001
- Hart, S. D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology, 3*, 121-137.
doi: 10.1111/j.2044-8333.1998.tb00354.x
- Hart, S. D., & Cooke, D. J. (2013). Another look at the (Im-) Precision of individual risk estimates made using actuarial risk assessment instruments. *Behavioral Sciences & the Law, 31*, 81-102. doi: 10.1002/bsl.2049
- Hart, S. D., Hare, R. D., & Cox, D. N. (1995). *The Hare Psychopathy Checklist: Screening Version (PCL: SV)*. Toronto, Ontario, Canada: Multi-Health Systems.
- Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: Evaluating the 'margins of error' of group v. individual predictions of violence. *The British Journal of Psychiatry, 190*, s60-s65. doi: 10.1192/bjp.190.5.s60
- Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior, 21*, 347-359. doi: 10.1023/A:1024851017947

- Heilbrun, K., O'Neill, M. L., Stevens, T. N., Strohman, L. K., Bowman, Q., & Lo, Y. W. (2004). Assessing normative approaches to communicating violence risk: A national survey of psychologists. *Behavioral Sciences & the Law*, *22*, 187-196. doi: 10.1002/bsl.570
- Heilbrun, K., O'Neill, M. L., Strohman, L. K., Bowman, Q., & Philipson, J. (2000). Expert approaches to communicating violence risk. *Law and Human Behavior*, *24*, 137- 148. doi: 10.1023/A:1005435005404
- Higgins, N., Watts, D., Bindman, J., Slade, M., & Thornicroft, G. (2005). Assessing violence risk in general adult psychiatry. *Psychiatric Bulletin*, *29*, 131-133. doi:10.1192/pb.29.4.131
- Hilton, N. Z., Carter, A. M., Harris, G. T., & Sharpe, A. J. (2008). Does using nonnumerical terms to describe risk aid violence risk communication? Clinician agreement and decision making. *Journal of Interpersonal Violence*, *23*, 171-188. doi: 10.1177/0886260507309337
- Hilton, N. Z., Harris, G. T., Rice, M. E., Houghton, R. E., & Eke, A. W. (2008). An indepth actuarial assessment for wife assault recidivism: The Domestic Violence Risk Appraisal Guide. *Law and Human Behavior*, *32*, 150-163. doi: 10.1007/s10979-007-9088-6
- Hilton, N. Z., Harris, G. T., Rice, M. E., Lang, C., Cormier, C. A., & Lines, K. J. (2004). A brief actuarial assessment for the prediction of wife assault recidivism: The Ontario Domestic Assault Risk Assessment. *Psychological Assessment*, *16*, 267-275. doi: 10.1037/1040-3590.16.3.267

- Hilton, N. Z., & Simmons, J. L. (2001). The influence of actuarial risk assessment in clinical judgments and tribunal decisions about mentally disordered offenders in maximum security. *Law and Human Behavior, 25*, 393-408.
doi: 10.1023/A:1010607719239
- Hoge, R. D., & Andrews, D. A. (2006). *Youth Level of Service/Case Management Inventory (YLS/CMI): User's Manual*. Toronto, Ontario, Canada: Multi-Health Systems.
- Khiroya, R., Weaver, T., & Maden, T. (2009). Use and perceived utility of structured violence risk assessments in English medium secure forensic units. *Psychiatric Bulletin, 33*, 129-132. doi:10.1192/pb.bp.108.019810
- Kroner, D. G., & Loza, W. (2001). Evidence for the efficacy of self-report in predicting nonviolent and violent criminal recidivism. *Journal of Interpersonal Violence, 16*, 168-177. doi: 10.1177/088626001016002005
- Kroner, D. G., & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior, 28*, 471-489. doi: 10.1177/009385480102800405
- Kroner, D., Mills, J., & Reddon, J. R. (2005). A coffee can, factor analysis, and prediction of antisocial behavior: The structure of criminal risk. *International Journal of Law & Psychiatry, 28*, 360-374. doi: 10.1016/j.ijlp.2004.01.011
- Kröner, C., Stadtland, C., Eidt, M., & Nedopil, N. (2007). The validity of the Violence Risk Appraisal Guide (VRAG) in predicting criminal recidivism. *Criminal Behaviour and Mental Health, 17*, 89-100. doi: 10.1002/cbm.644

- Kropp, R. P., Hart, S. D., Webster, C. D., & Eaves, D. (1999). *Spousal Risk Assessment Guide (SARA)*. Toronto, Ontario, Canada: Multi-Health Systems.
- Leistico, A. M. R., Salekin, R. T., DeCoster, J., & Rogers, R. (2008). A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. *Law and Human Behavior, 32*, 28-45. doi: 10.1007/s10979-007-9096-6
- Loza, W. (2005). *Self-Appraisal Questionnaire (SAQ): A tool for assessing violent and nonviolent recidivism*. Toronto, Ontario, Canada: Multi-Health Systems.
- Manchak, S. M., Skeem, J. L., & Douglas, K. S. (2008). Utility of the Revised Level of Service Inventory (LSI-R) in predicting recidivism after long-term incarceration. *Law and Human Behavior, 32*, 477-488. doi: 10.1007/s10979-007-9118-4
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, Minnesota, United States: University of Minnesota Press.
- Mills, J. F., & Gray, A. L. (2013). Two-tiered violence risk estimates: A validation study of an integrated-actuarial risk assessment instrument. *Psychological Services, 10*, 361-371. doi: 10.1037/a0032608
- Mills, J. F., & Kroner, D. G. (2005). *Two-tiered violence risk estimates*. Kingston, Ontario, Canada: Author.
- Mills, J. F., & Kroner, D. G. (2006a). The effect of discordance among violence and general recidivism risk estimates on predictive accuracy. *Criminal Behaviour and Mental Health, 16*, 155-166. doi: 10.1002/cbm.623

- Mills, J. F., & Kroner, D. G. (2006b). The effect of base-rate information on the perception of risk for reoffense. *American Journal of Forensic Psychology*, *24*, 45-56.
- Mills, J. F., Kroner, D. G., & Hemmati, T. (2007). The validity of violence risk estimates: An issue of item performance. *Psychological Services*, *4*, 1-12. doi: 10.1037/1541-1559.4.1.1
- Mills, J. F., Kroner, D. G., & Morgan, R. D. (2011). *Clinician's guide to violence risk assessment*. New York, New York, United States: Guilford Press.
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, *62*, 783-792. doi:10.1037/0022-006X.62.4.783
- Motiuk, L., & Nafekh, M. (2001). Using reintegration potential at intake to better identify safe release candidates. *Correctional Service of Canada: Forum on Corrections Research*, *13*, 11-13. Retrieved from http://www.csc-scc.gc.ca/research/forum/e131/131d_e.pdf
- Nafekh, M., & Motiuk, L. L. (2002). *The Statistical Information on Recidivism, Revised 1 (SIR-R1) Scale: A psychometric examination* (Report No. R126). Ottawa, Ontario, Canada: Correctional Service of Canada.
- Neumann, C. S., Hare, R. D., & Newman, J. P. (2007). The super-ordinate nature of the Psychopathy Checklist-Revised. *Journal of Personality Disorders*, *21*, 102-117. doi: 10.1521/pedi.2007.21.2.102

- Nuffield, J. (1982). *Parole decision-making in Canada: Research towards decision guidelines* (Report No. JS 22-65/1982E). Ottawa, Ontario, Canada: Solicitor General of Canada.
- Parent, G., Guay, J., & Knight, R. A. (2011). An assessment of long-term risk of recidivism by adult sex offenders: One size doesn't fit all. *Criminal Justice and Behavior, 38*, 188-209. doi: 10.1177/0093854810388238
- Quinsey, V. L., Book, A., & Skilling, T. A. (2004). A follow-up of deinstitutionalized men with intellectual disabilities and histories of antisocial behaviour. *Journal of Applied Research in Intellectual Disabilities, 17*, 243-253. doi: 10.1111/j.1468-3148.2004.00216.x
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington D.C., United States: APA.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk (2nd Ed.)*. Washington, D.C., United States: APA.
- Rice, M. E., & Harris, G. T. (1997). Cross-validation and extension of the Violence Risk Appraisal Guide for child molesters and rapists. *Law and Human Behavior, 21*, 231-241. doi: 10.1023/A:1024882430242
- 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. doi: 10.1007/s10979-005-6832-7
- Rice, M. E., Harris, G. T., & Lang, C. (2013). Validation and revision to the VRAG and SORAG: The Violence Risk Appraisal Guide-Revised (VRAG-R). *Psychological Assessment, 25*, 951-965. doi: 10.1037/a0032878

- Rice, M. E., Harris, G. T., Lang, C., & Bell, V. (1990). Recidivism among male insanity acquittees. *Journal of Psychiatry & Law, 18*, 379-403.
- Rorschach, H. (1942). *Psychodiagnostics*. Oxford, United Kingdom: Grune and Stratton.
- Scurich, N., & John, R. S. (2011). The effect of framing actuarial risk probabilities on involuntary civil commitment decisions. *Law and Human Behavior, 35*, 83-91. doi: 10.1007/s10979-010-9218-4
- Scurich, N., & John, R. S. (2012). Prescriptive approaches to communicating the risk of violence in actuarial risk assessment. *Psychology, Public Policy, and Law, 18*, 50-78. doi: 10.1037/a0024592
- Serin, R. C. (1991). *Development and validation of a psychological referral screening tool*. Ottawa, Ontario, Canada: Research and Statistics Branch, Correctional Service of Canada.
- Serin, R. C. (1996). Violent recidivism in criminal psychopaths. *Law and Human Behavior, 20*, 207-217. doi:10.1007/BF01499355
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428. doi: 10.1037/0033-2909.86.2.420
- Sjöstedt, G., & Grann, M. (2002). Risk assessment: What is being predicted by actuarial prediction instruments?. *International Journal of Forensic Mental Health, 1*, 179-183. doi: 10.1080/14999013.2002.10471172
- Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science, 20*, 38-42. doi: 10.1177/0963721410397271

- Skeem, J. L., Monahan, J., & Mulvey, E. P. (2002). Psychopathy, treatment involvement, and subsequent violence among civil psychiatric patients. *Law and Human Behavior, 26*, 577-603. doi:10.1023/A:1020993916404
- Skeem, J. L., & Mulvey, E. P. (2001). Psychopathy and community violence among civil psychiatric patients: Results from the MacArthur Violence Risk Assessment Study. *Journal of Consulting and Clinical Psychology, 69*, 358-374. doi: 10.1037/0022-006X.69.3.358
- Thornton, D. (2007). *Scoring guide for Risk Matrix 2000.9/SVC*. Birmingham, United Kingdom: Author.
- Tolman, A. O., & Mullendore, K. B. (2003). Risk evaluations for the courts: Is service quality a function of specialization?. *Professional Psychology: Research and Practice, 34*, 225-232. doi: 10.1037/0735-7028.34.3.225
- Viljoen, J. L., McLachlan, K., & Vincent, G. M. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment, 17*, 377-395. doi:10.1177/1073191109359587
- Vrana, G. C., Sroga, M., & Guzzo, L. (2008). *Predictive validity of the LSI-OR among a sample of adult male sexual assaulters (Unpublished manuscript)*. Nipissing University, North Bay, Ontario, Canada.
- Walters, G. D. (2003). Predicting institutional adjustment and recidivism with the Psychopathy Checklist factor scores: A meta-analysis. *Law and Human Behavior, 27*, 541-558. doi: 10.1023/A:1025490207678
- Walters, G. D. (2015). A two-dimensional model of psychopathy and antisocial behavior: A multi-sample investigation using items from the Psychopathy Checklist-

Revised. *Personality and Individual Differences*, 78, 88-93.

doi:10.1016/j.paid.2015.01.037

Walters, G. D., & Heilbrun, K. (2010). Violence risk assessment and Facet 4 of the Psychopathy Checklist: Predicting institutional and community aggression in two forensic samples. *Assessment*, 17, 259-268. doi:10.1177/1073191109356685

Walters, G. D., Knight, R. A., Grann, M., & Dahle, K. P. (2008). Incremental validity of the Psychopathy Checklist facet scores: Predicting release outcome in six samples. *Journal of Abnormal Psychology*, 117, 396-405. doi: 10.1037/0021-843X.117.2.396

Walters, G. D., White, T. W., & Denney, D. (1991). The Lifestyle Criminality Screening Form: Preliminary data. *Criminal Justice and Behavior*, 18, 406–418.

doi:10.1177/0093854891018004003

Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the Violence Risk Scale: A treatment-friendly violence risk assessment tool. *Psychology, Public Policy, and Law*, 12, 279–309. doi:10.1037/1076-8971.12.3.279

Woodworth, M., & Porter, S. (2002). In cold blood: Characteristics of criminal homicides as a function of psychopathy. *Journal of Abnormal Psychology*, 111, 436-444.

doi:10.1037/0021-843X.111.3.436

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva, Switzerland:

Author.

Wormith, J. S., & Goldstone, C. S. (1984). The clinical and statistical prediction of recidivism. *Criminal Justice and Behavior*, *11*, 3-34.

doi: 10.1177/0093854884011001001

Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior*, *39*, 1511-1538.

doi: 10.1177/0093854812455741

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

740-767. doi: 10.1037/a0020473

Appendices

Appendix A

Coding Manual

Section A: Identifying information

Offender Name: _____
FPS Number: _____ Research ID#: _____

Section B: Recidivism Data (Only for post-release from original assessment/index offense(s))

Release date from CSC post original index offense: ddmmyyyy

Type of Release: SRD = 1; DPD = 2; FPD = 3; WED = 4

Recidivism - Any Offense: No recidivism = 0; Recidivism = 1

Conviction Date - Any Offense First Incident, Only: ddmmyyyy

Recidivism - Type of Offense (from Hare PCL-R Item 20 and the Cormier-Lang Scales).

Type of Crime	# of Non-Violent	# of Violent
Property (Theft, B & E, possession of house breaking tools, possession of stolen property, loitering at night) = 100;		
Robbery (Robbery, extortion, etc) = 200;		
Armed Robbery (Armed robbery, robbery with violence, etc) = 201;		
Drugs (Possession, trafficking) = 300;		
Assault (Assault – simple, threatening, threatening with a weapon) = 400;		

Assault (Assault causing bodily harm, aggravated assault, assault with a weapon choking, administering a noxious substance, etc.) = 401;		
Assaulting a peace officer = 402;		
Murder (Murder, attempted murder, manslaughter, etc.) = 500;		
Weapons (possession of weapon, explosives) = 600		
Sexual (Sexual assault, aggravated sexual assault, gross indecency, etc) = 700;		
Indecent exposure = 701;		
Criminal negligence (Impaired driving, dangerous driving, hit and run, etc.) = 800		
Fraud (Fraud, forgery, false pretences, impersonation, uttering, etc.) = 900;		
Escape (Escape, UAL, jumping bail, FTA, breach of recognizance, etc.) = 1000;		
Kidnapping (Kidnapping, unlawful confinement, forcible seizure, hijacking, abduction, etc) = 1100;		
Arson = 1200		
Obstruction (Obstruction of justice, perjury, etc) = 1300		
Crimes against the state (Treason, espionage, smuggling, tax evasion, etc.) = 1400;		

Miscellaneous: (minor charges – for example, vandalism, casus a disturbance, mischief, willful damage, minor driving offenses, living off the avails, procuring, etc.) = 1500		
Total Violent & Non-violent		

Recidivism - Violent Offense: No recidivism = 0; Recidivism = 1

Conviction Date - Violent Offense First Incident, Only: ddmmyyyy

Recidivism – Non-violent Offense: No recidivism = 0; Recidivism = 1

Conviction Date – Non-violent Offense First Incident, Only: ddmmyyyy

Total Number of Violent Convictions Post Release from CSC:

Total Number of Non-Violent Convictions Post Release from CSC:

Total Number of Violent Convictions Overall:

Total Number of Non-Violent Convictions Overall:

Section C: Two-Tiered Violence Risk Estimates Scale (TTV)

Part A: Actuarial Risk Estimate

A1. Childhood antisocial behaviour

Childhood antisocial behaviour not present = 0

Childhood antisocial behaviour present = 1

A2. Adolescent antisocial behaviour

No adolescent antisocial behaviour = 0

Adolescent antisocial behaviour = 1

A3. Age of first adult conviction

Age 31 or more = 0

Age 20-30 = 1

Age 18-19 = 2

A4. Prior incarcerations

No prior incarcerations	= 0
1-3 prior incarcerations	= 1
4+ prior incarcerations	= 2

A5. Prior convictions for assaultive behavior

No prior assaultive convictions	= 0
1-3 prior assaultive convictions	= 1
4+ prior assaultive convictions	= 2

A6. Community supervision failure

No community supervision failure	= 0
Prior community supervision failure	= 1

A7. History of alcohol abuse

No history of alcohol abuse	= 0
History of alcohol abuse	= 1

A8. Failure to complete high school

Completed high school	= 0
Did not complete high school	= 1

A9. Criminal associations

No criminal associations	= 0
Criminal associations	= 1

A10. Interpersonal difficulties

No relationship difficulties	= 0
Relationship difficulties	= 1

Part B: Risk Management Indicators

M1. Employment	Not present	= 0
	Present and requires monitoring/assistance	= 1
	Present and requires intervention	= 2

M2. FINANCIAL	Not present	= 0
	Present and requires monitoring/assistance	= 1
	Present and requires intervention	= 2

M3. Substance abuse	Not present	= 0
	Present and requires monitoring/assistance	= 1
	Present and requires intervention	= 2
M4. Mental health	Not present	= 0
	Present and requires monitoring/assistance	= 1
	Present and requires intervention	= 2
M5. Family instability		
	Not present	= 0
	Present and requires monitoring/assistance	= 1
	Present and requires intervention	= 2
M6. Associates	Not present	= 0
	Present and requires monitoring/assistance	= 1
	Present and requires intervention	= 2
M7. Attitudes	Not present	= 0
	Present and requires monitoring/assistance	= 1
	Present and requires intervention	= 2
M8. Leisure	Not present	= 0

Present and requires monitoring/assistance = 1
Present and requires intervention = 2

M9. Resistance to intervention

Not present = 0
Present and requires monitoring/assistance = 1
Present and requires intervention = 2

M10. Mood

Not present = 0
Present and requires monitoring/assistance = 1
Present and requires intervention = 2

M11. Social Support

Not present = 0
Present and requires monitoring/assistance = 1
Present and requires intervention = 2

M12. Environment

Not present = 0
Present and requires monitoring/assistance = 1
Present and requires intervention = 2

M13. Stressors

Not present = 0
Present and requires monitoring/assistance = 1
Present and requires intervention = 2

Appendix B

Revised RMI Coding Manual

M1. Employment

Not present (0): Individual has work offer upon release, will be supported by disability pension, or is genuinely unable to work.

Present and requires monitoring (1): Individual is not employed but will be actively seeking employment upon release or has some concrete employment goals upon release.

Present and requires intervention (2): Individual is unwilling to seek employment upon release or is planning to earn money through criminal means.

M2. Financial

Not present (0): Individual has pension or ample means of support that is not crime-related (e.g., savings).

Present and requires monitoring (1): Individual has no means of financial support while in community and would have difficulty meeting financial obligations through legitimate means.

Present and requires intervention (2): Individual has no source of income, relies heavily on others for financial support, or supports self mainly through criminal activity.

M3. Substance Abuse

Not present (0): Individual has no history of substance abuse.

Present and requires monitoring (1): Individual has history of substance abuse but is currently not using substances and requires relapse prevention support.

Present and requires intervention (2): Individual has history of substance abuse and is currently still abusing substances.

M4. Mental Health

Not present (0): Individual has no history of or current symptoms of mental illness.

Present and requires monitoring (1): Individual has history of mental illness or is currently in the care of a psychologist, psychiatrist, or physician and is receiving intervention.

Present and requires intervention (2): Individual has mental health issues that are currently impeding their day-to-day functioning.

M5. Family Instability

Not present (0): Individual has a family situation that is supportive and free of ongoing conflict.

Present and requires monitoring (1): Individual is involved in a family relationship that is not supportive or where there is ongoing conflict.

Present and requires intervention (2): Individual is involved in a family relationship that is not supportive and where there is daily conflict that involves physical assault or threats.

M6. Associates

Not present (0): Individual has no associates that have a criminal record or are involved in criminal behaviour.

Present and requires monitoring (1): Individual associates with anyone who has a criminal record or who is involved with criminal activity, including family members.

Present and requires intervention (2): Individual associates with more than one person who participates in criminal behaviour.

M7. Attitudes

Not present (0): Individual holds prosocial views and attitudes.

Present and requires monitoring (1): Individual endorses antisocial behaviour, including the use of violence if provoked.

Present and requires intervention (2): Individual is openly antagonistic toward the law and social convention, and identifies criminal activity as a legitimate means of earning a living.

M8. Leisure

Not present (0): Individual participates in structured activities or spends time actively parenting or in the presence of prosocial family members while out in the community.

Present and requires monitoring (1): Individual spends leisure time in an unstructured and unproductive manner, including substance abuse, hanging out with friends, or watching TV.

Present and requires intervention (2): Individual has excessive free time (e.g., is unemployed) and does not have productive pursuits.

M9. Resistance to Intervention

Not present (0): Individual has participated in interventions as directed and expresses a willingness to comply with conditions of release.

Present and requires monitoring (1): Individual has not participated in recommended interventions or is not willing to comply with conditions of release.

Present and requires intervention (2): Individual has not participated in recommended interventions or is not willing to comply with conditions of release, and is openly antagonistic toward interventions.

M10. Mood

Not present (0): Individual has generally stable mood.

Present and requires monitoring (1): Individual has a deteriorating mood, with more prominent anger, sadness, or anxiety.

Present and requires intervention (2): Individual has a deteriorating mood, with more prominent anger, sadness, or anxiety that are strong and of immediate concern.

M11. Social Support

Not present (0): Individual has a social support system to assist them in the community.

Present and requires monitoring (1): Individual has no social support system in community.

Present and requires intervention (2): Individual has no social support system in community and their needs are extreme.

M12. Environment

Not present (0): Individual is not returning to a neighbourhood that is high in crime or is related to their criminal offending.

Present and requires monitoring (1): Individual is returning to a high-crime neighbourhood or one that is associated with previous offending.

Present and requires intervention (2): Individual is returning to a high-crime neighbourhood or one that is associated with previous offending. This neighbourhood also has few opportunities for professional services or employment opportunities.

M13. Stressors

Not present (0): Individual is able to cope with day-to-day stressors and no extreme stressors are present.

Present and requires monitoring (1): Stressors are present and individual is having problems with coping.

Present and requires intervention (2): Stressors are present and individual is overwhelmed by them.

Appendix C

Items of the Violence Risk Appraisal Guide (VRAG) and the Statistical Information on Recidivism (SIR-R1)

VRAG

1. Lived with both parents to the age of 16
2. Elementary school maladjustment
3. History of alcohol problems
4. Marital status
5. Criminal history score using non-violent offenses
6. Failure on conditional release
7. Age at index offense
8. Victim injury
9. Female victim
10. DSM diagnosis of any personality disorder
11. DSM diagnosis of schizophrenia
12. PCL-R or CATS score

SIR-R1

1. Current offense
2. Age at admission
3. Previous incarceration
4. Revocation or forfeiture
5. Act of escape
6. Security classification
7. Age at first adult conviction
8. Previous convictions for assault
9. Marital status at most recent admission
10. Interval at risk since last offense
11. Number of dependents at most recent admission
12. Current aggregate total sentence
13. Previous offenses for sex offenses
14. Previous convictions for breaking and entering
15. Employment status at arrest

Appendix D

Correlations Between Study Measures

	PCL-R total	PCL-R F1	PCL-R F2	VRAG	SIR-R1
PCL-R total	-	.75***	.83***	.68***	.54***
PCL-R F1		-	.31***	.23*	.12
PCL-R F2			-	.84***	.70***
VRAG				-	.67***
SIR-R1					-

Note. PCL-R = Psychopathy Checklist – Revised; F1 = Factor 1; F2 = Factor 2; VRAG = Violence Risk Appraisal Guide; SIR-R1 = Statistical Information on Recidivism – Revised.

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