

**Assessing the Utility of
Risk Assessment Tools and
Personality Measures in the
Prediction of Violent Recidivism
for Adult Offenders**

2007-04

by
Mary Ann Campbell, Ph.D.
Sheila French, M.A.
and
Paul Gendreau, Ph.D.
Centre for Criminal Justice Studies
University of New Brunswick-Saint John

This report was prepared under contract to the Department of Public Safety and Emergency Preparedness. Views expressed herein are those of the authors and do not necessarily reflect those of the Department.

Cat. No.: PS3-1/2007-4E-PDF
ISBN No.: 978-0-662-46071-8

Authors' Note

The first author is Assistant Professor and Director of the Centre for Criminal Justice Studies, Department of Psychology, University of New Brunswick (Saint John campus). The second author is a Psychology Doctoral Student, supervised by the first author. The third author is a Visiting Scholar with the Division of Criminal Justice at the University of Cincinnati and Professor Emeritus in the Department of Psychology at the University of New Brunswick (Saint John campus). Further inquiries about the manuscript should be forwarded to Mary Ann Campbell, Ph.D. at the Centre for Criminal Justice Studies, University of New Brunswick, PO Box 5050, Saint John, New Brunswick, Canada, E2L 4L5; Email: mcampbel@unbsj.ca.

The authors would like to recognize the assistance of Laurie Green, who conducted the inter-rater coding for this project. We are grateful to Karl Hanson for suggestions with respect to base rate consideration and to Paula Smith for assistance with database construction. Finally, we appreciate the time and effort spent by Delphine Gossner, Karl Hanson, Peter Raynor, Steve Van Dine, Glenn Walters, and Steve Wormith, who so graciously offered references and/or unpublished data.

An earlier version of this report was released in 2007, but has since been revised due to statistical errors. Little to no changes resulted in the interpretation, conclusion, or recommendations of that report, but citations should refer to the current document for clarity and accuracy.

Table of Contents

Executive Summary	i
Introduction.....	1
Predictors of Risk and Procedural Principles Important to Risk Assessment	1
Variations in Violence Risk Assessment Instruments.....	2
Summary	6
Methods and Procedures	7
Sample of Studies.....	7
Coding of Studies.....	7
Effect Size Calculation.....	7
Effect Size Heterogeneity	8
Fail Safe Estimation	8
Results.....	11
Description of Database	11
Risk Measures: Predictive Validities for Institutional Violence	12
Risk Measures: Predictive Validities for Violent Recidivism.....	12
Comparison of Effect Sizes by Generation of Risk Instrument	13
Comparisons Based on the Content of the Instrument: Static versus Dynamic	14
Comparisons Based on Measure Administration Method.....	15
Comparisons Based on Instrument Relevance to Corrections	16
Discussion	19
Actuarial, Structured, and Psychopathy Checklist Instruments	19
Recommendations to Guide Instrument Selection for the Assessment of Future Violence.....	22
Further Issues for Future Research and Consideration	26
Conclusion	27
References.....	29
Appendix A.....	43

Executive Summary

Despite the availability of violence-specific risk assessment tools (e.g., VRAG), general risk instruments associated with persistent criminality and violence (e.g., LSI-R), and personality-based measures associated with aggression (e.g., PCL-R), relatively few meta-analytic comparisons of these assessment approaches have been conducted with regard to their predictive validity and ultimate value to violence risk assessment objectives. These objectives include the prediction of risk, the identification of risk reduction targets, and the provision of a means to monitor changes in risk level. Thus, the objective of the current study was to conduct a meta-analytic evaluation of the relative utility of risk instruments and other psychological measures as a means of informing the standards of practice for conducting violent risk assessments.

The current meta-analysis compared various instruments (e.g., self-report, actuarial, structured clinical risk protocols) that have been used to assist with the estimation of violence risk. In order to be included in the meta-analysis, a study had to be conducted after 1980, be prospective in nature, and involve an adult offender/forensic psychiatric sample. Also essential for inclusion was that the study reported a statistical estimate of the relationship between a particular assessment instrument and a violent (non-sexual) outcome relating to either institutional violence or violent recidivism that could be converted to an effect size. Recidivism studies also must have included a post-release follow-up period of at least six months to be considered for analysis, while no minimum follow-up period was imposed for studies focusing on institutional violence as long as it was prospective in nature. Every effort was made to gather both published and unpublished data.

A coding guide was developed to obtain information pertaining to study and sample characteristics, details of the type and format of the risk assessment methods used, and to collect the statistical information relevant to the calculation of effect size estimates. Interpretation of the results was based on the Z^+ statistic, which is an effect size estimate that has been adjusted for sample size. The Z^+ statistic itself was calculated using r' , which is the correlation between an instrument and an outcome variable after being adjusted for the violence base rate of the sample. When the 95% confidence intervals between any two mean effect size comparisons overlapped by no more than one quarter of the average length of the two intervals, the two mean effects were interpreted as representing two different population parameters and were statistically different from each other (i.e., $p \leq .05$; Cummings & Finch, 2005). Overlap in effect size confidence intervals that exceeded this criterion indicated that the effects were produced from the same sample population parameter, and therefore, not statistically different from each other.

A number of meaningful findings were obtained from the current meta-analysis. Most notably, there was very little variation in the magnitude of the predictive validities for violent recidivism amongst the more commonly used and researched actuarial/structured risk instruments (i.e., HCR-20, LSI-R, VRAG, SIR, and PCL-R). The greatest number of effect sizes was obtained for the PCL-R, but all five instruments produced relatively precise confidence intervals for their estimates of violence risk in the community. The effect sizes ranged from .22 (for the SIR and HCR-20 scales) to .32 (for the VRAG), which reflect an overall moderate predictive validity of these instruments for the outcome of violence recidivism. The confidence intervals for the VRAG, LSI-R, and PCL-R all overlapped, indicating that these measures were equally predictive of violent recidivism. The effect size confidence intervals for the HCR-20 also overlapped with the LSI-R, PCL-R, and SIR scale, while the VRAG performed somewhat more proficiently than the HCR-20 and the SIR scale.

Substantially fewer effect sizes were available for the prediction of institutional violence than had been obtained for violent recidivism. Only the HCR-20 exceeded the minimum criteria of 10 effect sizes for meaningful interpretation. A great deal of variability was also evident in the confidence intervals for the effect sizes of each instrument examined (including the HCR-20), which further limited the interpretation of their predictive validities. With these limitations in mind, preliminary data suggest that the strongest predictive estimates came from the HCR-20 ($Z^+ = .28$), the LSI-R ($Z^+ = .24$), the PCL-SV ($Z^+ = .22$), and a miscellaneous criminal history variable ($Z^+ = .26$). Although only the HCR-20 had enough effect sizes to allow relative confidence in the interpretation of its true predictive validity, these effects were primarily based on forensic psychiatric samples. Thus, even the use of the HCR-20 as a predictor of institutional violence requires additional research regarding its applicability to general offender institutional settings.

Other variables were examined as possible moderators of the predictive validities for the instruments used to assess violence risk. Based on these analyses, as a group, second generation risk instruments (developed from statistical procedures, contained primarily static items, and were atheoretical in nature) were the strongest predictors of institutional violence. In contrast, third generation risk instruments (theoretically based, inclusion of dynamic items, and concerned with measuring changes in risk) collectively provided the strongest effect sizes for the prediction of violent recidivism in the community. The variation in predictive validity of second and third generation instruments across violent recidivism and institutional violence outcomes may be due to the length of follow-up periods. Institutional violence studies tend to be shorter in duration relative to recidivism studies. Further, instruments that contained content relevant to criminological theory and/or those specifically designed as risk assessment instruments were more strongly tied to predictions of institutional violence and violence recidivism than were those not originally designed for risk prediction purposes and/or that measured irrelevant or weakly associated constructs to criminal behaviour (e.g., self-esteem).

In conclusion, most of the currently available risk instruments are moderately predictive of future violence. Instruments based on historical factors may provide more reliable estimates of the violence risk during incarceration (i.e., short-term predictions), but the inclusion of dynamic risk-need factors is important to the prediction of violence once released to the community (i.e., long-term predictions). Given that most of these instruments predict violence with similar degrees of accuracy, the selection of the most appropriate risk instrument should be based on: (a) the purpose of the risk assessment; (b) the instrument's ability to adequately identify relevant criminogenic needs that contribute to an offender's risk of violence; (c) the instrument's informative value for treatment programming to reduce the risk of violence; and (d) the instrument's capacity to measure changes in risk level. Each of these factors facilitates effective case management practices for offenders in custody and under supervision in the community.

Assessing the Utility of Risk Assessment Tools and Personality Measures in the Prediction of Violent Recidivism for Adult Offenders

Assessments of an offender's risk of future violence play a central role in decision-making pertaining to that individual's sentencing, case management, community release, and public safety concerns (Andrews & Bonta, 2003; Hoge & Andrews, 1996). These assessments should also help guide the selection of intervention goals and strategies that will lead to risk reduction (Heilbrun, 1997). Much of the current knowledge-base regarding violence risk prediction was accumulated in response to concerns (primarily in the 1950s and 1960s) over the validity of the criteria being used to render risk decisions (c.f. Andrews, 1989; Heilbrun, 1997; Litwack & Schlesinger, 1999; Monahan & Steadman, 1994; Rice, 1997). Specifically, the first generation of risk assessments arising in the mid-20th century (see Bonta, 2002) was based on subjective clinical judgments of risk, which were typically formulated using unstructured and unsystematic assessment methods (Hoge & Andrews, 1996). For the most part, subsequent research has shown that the accuracy of unstructured risk judgments is inferior to estimates of risk derived from objective, structured, and evidence-based (actuarial) methods of prediction (Bonta, Law, & Hanson, 1998; Grove, Zald, Lebow, Snitz, & Nelson, 2000). The bottom line is that accurate prediction of future dangerousness has proven to be a difficult task for professionals (Hanson, 2005; Quinsey, Harris, Rice, & Cormier, 1998), but can be facilitated by the use of structured risk assessment instruments. To further facilitate the practice of violence risk assessment, the current study will provide a meta-analytic review of the various assessment tools that have been used to guide and inform this process. To set a foundation for the results and discussion, the first section of this introduction will briefly outline the factors of import for accurate risk prediction, as well as the principles that should guide the risk assessment/reduction process. This will be followed by a review of contemporary violence risk assessment tools and their variations in terms of their predicted outcome, content, and administration format.

Predictors of Risk and Procedural Principles Important to Risk Assessment

A necessary feature of effective risk assessment is the identification of variables contributing to and sustaining an individual's involvement in criminal behaviour (Bonta, 2002). Much research has been dedicated to this task and has highlighted a number of important historical and psychosocial factors as relevant to the prediction of dangerousness and persistent criminality (see Andrews & Bonta, 2003; Borum, 1996). To briefly summarize two comprehensive meta-analyses on the predictors of recidivism, Bonta et al. (1998) and Gendreau, Little, and Goggin (1996), have identified antisocial attitudes, antisocial peer associates, substance abuse, family dysfunction, interpersonal conflict, and negative/unstable living arrangements, as well as some demographic variables (being male, single, and younger in age), as useful predictors of general and violent recidivism. With specific reference to violent recidivism, additional predictors include a diagnosis of Antisocial Personality/Psychopathy, a history of violent behaviour, and employment problems (Bonta et al., 1998). Further, these major predictors of risk are common to both general offender populations and mentally disordered offenders (Bonta et al., 1998; Phillips et al., 2005). It is also worthwhile to note that many of the traditional clinical factors used for judgments of risk (e.g., intelligence, mood disorders, psychosis, self-esteem) have produced the smallest predictive validities for violent and general recidivism (Bonta et al., 1998; Gendreau et al., 1996).

Accurate assessment of risk is an essential step in the successful reduction of risk. The guiding principles underlying efficient risk assessment/rehabilitation are the risk, need, and responsivity principles described

by Andrews and Bonta (2003). According to their model, the *risk principle* is based on the premise that criminal behaviour can be predicted and that the intensity of intervention to reduce this risk should be matched to the offender's risk level. Secondly, the *need principle* recognizes that certain risk factors are capable of being changed in a manner that reduces risk. These "criminogenic needs" relate to an offender's lifestyle, cognitions, and behaviour (e.g., antisocial attitudes, substance abuse) and are empirically tied to the risk of violence and/or general criminality. As such, interventions designed to reduce risk should address criminogenic needs, rather than factors that have weak ties to recidivism (e.g., self-esteem, depression). Lastly, the *responsivity principle* is concerned with the style and method of intervention used to target criminogenic needs. Essentially, the choice of treatment should be based upon empirically-supported programs for the reduction of criminal behaviour, such as cognitive-behavioural and social learning approaches (i.e., general responsivity; Andrews & Bonta, 2003). The intervention also should be sensitive to the offender's learning style and other factors that may interfere with his or her ability to respond to the intervention, such as mental disorder, motivation to change, or physical impairments (i.e., specific responsivity; Andrews & Bonta, 2003). Adherence to the risk-need-responsivity principles has been shown to contribute to greater risk reduction than interventions ignoring or minimally adopting these principles (Andrews & Bonta, 2003; Dowden & Andrews, 2000; French & Gendreau, 2006; Gendreau, Goggin, French, & Smith, 2006). The extensive research on the prediction of risk and the risk-need-responsivity principles has provided meaningful guideposts from which to construct valid instruments for the purposes of violence risk assessment.

Variations in Violence Risk Assessment Instruments

With the goal of improving the quality, validity, and efficiency of violence risk decision-making, much attention has been dedicated to the development of standardized risk assessment tools and a number of promising empirically-derived risk instruments have been developed (Borum, 1996; Webster, Douglas, Eaves, & Hart, 1997a). Some of these instruments are specifically designed to predict dangerousness, such as the Violence Prediction Scheme (VPS; Webster, Harris, Rice, Cormier, & Quinsey, 1994), the Violence Risk Scale (VRS; Wong & Gordon, 2006), and the Historical, Clinical, and Risk Management Violence Risk Assessment Scheme (HCR-20; Webster, Douglas, Eaves, & Hart, 1997b). In a few cases, instruments are meant to predict a specific form of violence, such as intimate partner violence (Spousal Assault Risk Assessment Guide - SARA; Kropp, Hart, Webster, & Eaves, 1995) or sexual recidivism (Sexual Violence Risk-20 - SVR-20; Boer, Hart, Kropp, & Webster, 1997). Hence, professionals have a choice of assessment tools for the prediction of general dangerousness, as well as certain types of violent behaviour.

Although not developed as a risk measure, psychopathic personality traits as assessed by the Psychopathy Checklist-Revised (PCL-R; Hare, 1991; 2003) have proven useful in the prediction of future violence (Gendreau, Goggin, & Smith, 2002; Hemphill, Hare, & Wong, 1998). In addition, general recidivism risk measures (e.g., Level of Supervision Inventory-Revised, LSI-R; Andrews & Bonta, 1995) have been shown to predict future violence with reasonable success (Gendreau et al., 2002; Harris, Rice, & Quinsey, 1993). The utility of general risk instruments for predicting violence is likely due to the overlap in risk predictors for violent and general recidivism (Bonta et al., 1998). Thus, in addition to violence-specific instruments, measures designed for other purposes may assist in the prediction of violence.

As described by Bonta (2002), potential formats for the assessment of risk include paper-and-pencil methods (e.g., the Criminal Sentiments Scale – CSS; Andrews & Wormith, 1984; the Self-Appraisal Questionnaire – SAQ; Loza, Dhaliwal, Kroner, & Loza-Fanous, 2000), file review methods (e.g.,

Violence Risk Appraisal Guide - VRAG; Harris et al., 1993), and interview-based approaches that are combined with file reviews (e.g., LSI-R, HCR-20). Some of these approaches measure a single construct relevant to risk (e.g., antisocial attitudes as measured by the modified CSS, Simourd, 1997), while others tap multiple domains associated with recidivism (e.g., the LSI-R assesses 10 risk-need domains). Although there is diversity in both the administration format and content areas of risk assessment tools, the practice of combining risk instruments to generate a consensus estimation of risk can be problematic. Mills and Kroner (2006) used the PCL-R, LSI-R, VRAG, and General Statistical Information of Recidivism (GSIR; Bonta, Harman, Hann, & Cormier, 1996; Nuffield, 1982) to predict post-release violent and general recidivism. For most offenders, there was agreement in the standardized risk scores generated for each of these instruments. Unfortunately, predictive accuracy was substantially reduced for cases in which there was a high level of disagreement between instruments in their standardized risk scores. The challenges associated with formulating risk judgments based on the use of several risk instruments highlight the need for research that identifies the most appropriate risk instrument for a given offender population, forensic setting, and assessment purpose.

Generations of Risk Instruments. The nature of the information contained within risk instruments varies in its usefulness for rehabilitation objectives and for the monitoring of changes in risk over time. As mentioned above, first generation risk instruments were based on unstructured, non-systematic, and subjective clinical judgments of risk and were prone to error and bias (e.g., Grove et al., 2000). In light of the limitations with first generation methods of assessment, the second generation of risk instruments were designed to provide proficient, standardized risk predictions. However, the item selection for second generation methods was purely statistically driven (i.e., actuarial; Bonta, 2002). Only items that were maximally predictive of recidivism were included (e.g., being male, ethnicity), regardless of their theoretical or rehabilitative value. Examples are the VRAG (Harris et al., 1993), the Salient Factor Score (SFS; Hoffman, 1983), and the GSIR (Bonta et al., 1996; Nuffield, 1982). Despite the fact that some actuarial risk instruments demonstrate fairly good predictive validities (e.g., .30 to .35; Bonta & Yessine, 2005; Glover, Nicholson, Hemmati, Berfeld, & Quinsey, 2002; Loza & Green, 2003; Polvi, 2001), they are predominately composed of static items (Andrews & Bonta, 2003). Static risk factors are historical in nature and/or are unchangeable (e.g., gender, age, history of prior offences). Among static factors, criminal history is one of the stronger predictors of future violence and general recidivism (Bonta et al., 1998; Gendreau et al., 1996; Webster et al., 1997a). However, predictors based on previous offences have been criticized because they do not capture the complexity of factors contributing to recidivism, nor do they measure change in risk level over time, and both are essential for case management to reduce risk (Andrews, Bonta, & Hoge, 1990; Hoge & Andrews, 1996). Critics of second generation instruments argue that the true goal of risk assessment should be to inform risk reduction, rather than to solely predict risk (Wong & Gordon, 2006).

In response to these criticisms, the third generation of risk instruments (Andrews, Bonta, & Wormith, 2006; Bonta, 2002) emphasized the informative value of prediction models for case management decision-makers (e.g., parole/probation officers, parole boards, forensic psychologists). As with second generation measures, these instruments included empirically-supported risk factors; however, item selection was driven by theoretical understandings of persistent criminality and violence (i.e., social learning and social cognition theories, the principles of risk-need-responsivity; Andrews & Bonta, 2003; Gendreau et al., 2006). Moreover, third generation risk measures included dynamic risk factors, which are variable in nature and can change with time or with the influence of social, psychological, biological, or contextual factors, such as treatment intervention. Examples of such malleable factors are substance use,

interpersonal conflict, and antisocial attitudes (see Douglas & Skeem, 2005). Although static and dynamic risk factors have proven to be equally useful in predicting risk (Gendreau et al., 1996; Wong & Gordon, 2006), current theory advocates that dynamic factors are more relevant when the focus is on risk reduction (Andrews, 1989; Douglas & Skeem, 2005; Heilbrun, 1997). Within the context of risk assessment, dynamic factors are often referred to as “criminogenic needs” because of their empirical ties to criminal behaviour. Thus, the advantage of using instruments that assess dynamic risk factors is that they are sensitive to changes in risk level that might occur over time and/or as a result of rehabilitation efforts (Andrews & Bonta, 2003; Heilbrun, 1997).

The most recent generation of risk instruments (i.e., fourth generation; Andrews & Bonta, 2003; Andrews, et al., 2006) has been specifically designed to be integrated into: (a) the process of risk management; (b) the selection of intervention modes and targets for treatment; and (c) the assessment of treatment progress. These instruments are administered on multiple occasions (i.e., re-assessments) and are informative because they document changes in specific criminogenic needs, and in the overall risk potential, that might occur between an offender’s initial contact with the criminal justice system through to his or her exit from the system. As such, fourth generation instruments can identify areas of success within a case management plan designed to reduce risk, as well as identify areas where strategies should be modified to maximize their potential for reducing risk. Dynamic forms of risk assessment instruments are not very common at present (Bonta, 2002), but two promising examples are the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004) and the Violence Risk Scale (VRS; Wong & Gordon, 2006).

Clearly, the practice of predicting future dangerousness and criminality has greatly improved since the use of unstructured clinical risk judgments. However, there is still inconsistency with regard to whether, and to what degree, clinical judgment should be incorporated into the risk assessment procedure. Although the objective of second generation instruments was to leave little to no room for subjective clinical judgment, some risk measures encourage a degree of clinical flexibility in the rendering of risk estimates (e.g., LS/CMI, HCR-20, SARA). For example, several instruments (e.g., LS/CMI) allow the assessor to use a clinical “over-ride,” which means that the actuarially-derived estimate of risk can be adjusted based on the assessor’s subjective judgments about the role of protective factors, mitigating circumstances, or other factors unique to a case. Clinical flexibility is also contained in instruments that discourage the use of purely actuarial computations of risk estimates when used for clinical purposes. For example, there are no clinical cut-off scores or numerically-based risk probabilities for the HCR-20 (Webster et al., 1997b). Instead, the risk estimate is based on the assessor’s subjective judgment as to whether the offender falls within a “low,” “moderate,” or “high” risk level category. This judgment is based on a systematic and careful review of theoretically and empirically-relevant risk factors identified within the HCR-20 scheme. To reflect the structure contained within the review of individually rated risk factors, this approach to risk assessment is referred to as “structured prediction judgment.” However, instruments based on this approach have been criticized for being overly subjective and, as such, open to some of the same limitations found within unstructured clinical risk judgments (e.g., Hilton, Harris, & Rice, 2006).

Administration Format and Content Relevance of Risk Assessment Measures. Another consideration in the choice of risk instrument is whether to use an observer-rated measure and/or a self-report measure. As described by Bonta (2002), potential formats for the assessment of risk include paper-and-pencil methods (e.g., CSS; Andrews & Wormith, 1984; Self-Appraisal Questionnaire – SAQ; Loza, Dhaliwal, Kroner, & Loza-Fanous, 2000), file review methods (e.g., VRAG; Harris et al., 1993), and interview-based

approaches that are combined with file reviews (e.g., LSI-R, HCR-20). The majority of risk prediction instruments are based on a trained professional's ratings of individual static and/or dynamic risk factors. These ratings are made after an extensive review of collateral and correctional file information, which may also include a semi-structured interview with the offender. This is a time consuming approach, but one that can yield a comprehensive and valid assessment of risk. Self-report measures (e.g., the MMPI-2 Pd scale: Hathaway & McKinley, 1967; the MMPI Megargee Classification System: Megargee & Bohn, 1979; and the Antisocial Features and Aggression subscales of the Personality Assessment Inventory: Morey, 1991) have also been utilized as a source of information for judgments about an offender's risk (e.g., Douglas, Hart, & Kropp, 2001; Magargee & Carbonell, 1995; Morey & Quigley, 2002; Osberg & Poland, 2001). Although self-report measures can be time and cost-efficient to administer, one major criticism is that they have not typically been designed to inform judgments of risk. As such, they are not necessarily representative of the empirically-identified risk-need factors relevant to meaningful risk prediction and management (Bonta, 2002; Walters, 2006).

In response to the criticism, some researchers have developed self-report instruments that are specifically designed to evaluate factors relevant to criminal risk outcomes. The Criminal Sentiments Scale-Modified (CSS-M; Simourd, 1997; Simourd & Van de Ven, 1999), the Psychological Inventory of Criminal Thinking Styles (PICTS; Walters, 1995 and 1996), and the Self-Appraisal Questionnaire (SAQ; Loza et al., 2000) are examples of risk-relevant self-report instruments. To test the value of self-report instruments in risk prediction, Walters (2006) conducted a meta-analysis that compared selected structured/actuarial risk instruments (i.e., HCR-20, LSI-R, PCL-R, VRAG, and the Lifestyle Criminality Screening Form – LCSF created by Walters, White, & Denney, 1991) with a number of self-report measures that have been used to inform risk judgments for institutional misconduct, general recidivism, and violence. Some of the self-report measures included were specific to risk prediction (e.g., PICTS, SAQ), while others reflected general clinical constructs thought to be relevant to risk, or at least to an individual's general personality and emotional functioning (e.g., NEO Personality Inventory-Revised, Multi-dimensional Anger Inventory, Beck Hopelessness Scale; MMPI-Pd scale). Walters's findings supported the predictive validity of self-report measures in risk assessment, but *only if* these instruments were based on constructs that are empirically-tied to risk (e.g., antisocial attitudes). Walters suggested that the integration of *content-relevant* self-report measures with actuarial/structured risk instruments could add to the validity of risk assessment.

Although an informative first step, there are several other avenues of interest that arise from Walter's (2006) meta-analysis. First, only a select number of structured/actuarial risk instruments were coded (HCR-20, LSI-R, PCL-R, VRAG, and LCFS). In addition, only nine effect sizes were available to compare the aggregate category of structured/actuarial methods with self-report measures in terms of their ability to predict *violent* recidivism. Across these nine effect sizes, the mean effect size for structured/actuarial measures ($r = .24$) was larger than it was for the general category of self-report measures ($r = .17$). A larger database, including more effect sizes and encompassing a greater range of measures, is required to replicate Walters' findings.

Walters (2006) also did not report the respective predictive validities of the individual instruments included within the actuarial/structured category. This type of information would prove valuable to professionals when deciding which of the available instruments they should incorporate into their violence risk assessments. A few meta-analyses have been conducted that relate to this issue. Gendreau et al. (1996) compared the LSI-R, the SFS, and the risk-need based Wisconsin Classification System (Baird, 1981, Baird, Heinz, & Bemus, 1979). Each of these instruments was moderately predictive of general

recidivism, but the LSI-R produced the strongest mean weighted effect size (.33). In addition, Gendreau et al. evaluated the value of the MMPI in predicting general recidivism. Although the MMPI was not as strong a predictor as the LSI-R or the PCL-R, it nonetheless produced a significant weighted effect size of .21. Unfortunately, Gendreau et al. did not analyze these instruments in relation to violent outcomes. An earlier meta-analysis by Gendreau, Goggin, and Law (1997) on the prediction of prison misconducts included a comparison of the LSI-R, MMPI, "other" risk measures, and non-MMPI measures of antisocial personality as predictors of institutional misconduct. As with Gendreau et al. (1996), the authors did not separately report the predictive validities of these measures for non-violent and violent misconducts. This decision was based on the lack of substantial variation in effect sizes across these two outcomes for the numerous other predictors examined in their analysis. Based on the aggregate outcome, the LSI-R produced the highest predictive validities ($r = .23$) and outperformed the other measures. Finally, a more recent meta-analytic comparison by Gendreau et al. (2002) did evaluate violence as a separate criterion and found that the LSI-R had a slight advantage over the PCL-R in the prediction of violence. Therefore, the existing meta-analyses that compare risk instruments suggest that there may be similarities in the predictive validity of risk instruments for violent and general recidivism.

Summary

Many advances have been made in the assessment of general risk and dangerousness. Nonetheless, uncertainty remains concerning the most appropriate instruments for the prediction of violence given variations in item content, scale format, level of permitted assessor subjectivity, and the utility of self-report instruments as a component of violence risk assessment protocols. Although several primary studies compare the utility of risk instruments for the prediction of violence (e.g., Dahle, 2006; Douglas, Yeomans, & Boer, 2005; Grann, Belfrage, & Tengstöm, 2000; Kroner & Mills, 2001; Mills & Kroner, 2006), only a few meta-analyses (i.e., Gendreau et al., 1996; Gendreau et al., 1997; Gendreau et al., 2002; Walters, 2006) have been conducted to synthesize this literature for professionals and none of these have been sufficiently comprehensive in their estimation of violence risk. A synthesis of this nature is timely given that very few correctional psychologists report using instruments specifically designed for risk prediction (or that are at least empirically-supported as relevant to the task of risk estimation, see Boothby & Clements, 2000). Thus, the primary objective of the current meta-analysis was to determine which instruments function most effectively as valid predictors of future violence (non-sexual) within prison settings and in the community. With this information, guidelines can be made regarding the selection of risk instruments with the potential to provide the most valid estimates of risk, as well as to inform case management and rehabilitation planning.

Methods and Procedure

Sample of Studies

An electronic literature search was conducted for relevant prediction studies via EBSCO databases (Academic Search Elite, PsycARTICLES, and PsycINFO). Key search terms included: (a) assessment-related terms (e.g., actuarial, clinical, prediction, LSI-R, PCL-R); (b) terms related to the offender population (e.g., adult offender, prisoner, parolee); and (c) terms meant to index the violent outcome (e.g., recidivism, misconduct). Unpublished data was also secured subsequent to an email request sent to approximately 33 researchers and 23 research centres known to conduct offender risk research. Additional studies were added via the ancestry method (i.e., review of article reference sections). The search was restricted to studies conducted from 1980 to 2006.

Inclusion criteria required that primary studies: (a) were predictive in nature (i.e., assessment preceded the measurement of outcome); (b) involved adults (i.e., sample mean of 18⁺ years at time of assessment) sampled from general or forensic offender populations; and (c) reported sufficient data to calculate an effect size (e.g., Pearson r , Phi coefficient Φ) between the prediction measure and violent misconduct/recidivism outcomes. Prison/probation studies were included regardless of length of follow-up. Post-release recidivism studies required at least a 6-month follow-up period for inclusion. For each study, data from the largest sample, longest follow-up period, and most specific type of criterion (i.e., conviction vs. arrest) was recorded. To avoid redundancy with Hanson and Morton-Bourgon's (2007) recent meta-analysis on the predictive validity of risk instruments for sexual offenders, the current analysis excluded studies for which violent outcome data was derived almost exclusively from a sex offender sample. Likewise, instruments designed specifically to assess sexual recidivism were not included in the current analysis. Included studies are detailed in Appendix A.

Coding of Studies

The coding categories, with examples of their sub-components, were as follows: (a) study/author characteristics (e.g., type of publication, author affiliation, publication year); (b) sample variables (e.g., ethnicity, gender, offender type); (c) risk assessment descriptors (e.g., measure used, administration method, type of predictors assessed); and (d) effect size descriptors (e.g., type of outcome, calculated effect size). Details or a copy of the coding manual can be obtained by contacting the first author. All studies were coded by S. French. Inter-rater reliability was established using a randomly selected sample of 15 studies, blindly coded by a second experienced coder. Using the Yeaton and Wortman (1993) formula: $\sum (\text{agreements}) / \sum (\text{agreements} + \text{disagreements})$, the index for agreement was .82. The source of disagreements concerned less obvious sample characteristics (i.e., determination of sample risk level) and aspects of the nature of a particular risk instrument (i.e., type of item content, generation of risk instrument). Disagreements most often resulted from a misunderstanding when reading the study or a clerical error when entering item codes. The two raters discussed disagreements and a consensus coding was achieved for those items prior to analysis.

Effect Size Calculation

Phi coefficients (Φ) were calculated for each measure's predictive validity with misconduct and recidivism outcomes. Where statistics other than r were reported (i.e., F , t , χ^2 , p , AUC), the appropriate

formula for conversion to Φ was employed (Rosenthal, 1991; Swets, 1986). In light of generally low base rates for violent misconduct and recidivism, it was necessary to consider this potential influence on effect sizes. Phi coefficients were adjusted using Ley's (1972) formula: $r' = [(r_{xy})(\delta_x'/\delta_x)] / [1 - r_{xy}^2 + (r_{xy}^2)(\delta_x'^2/\delta_x^2)]^{1/2}$, where r_{xy} was the observed correlation, δ_x was the observed standard deviation of the base rate, δ_x' was the average standard deviation based on the average base rate for studies in the analysis, and r_{xy}' was the corrected correlation. The standard deviation of the base rate was calculated using the formula: $\delta = [pq/(N)(N-1)]^{1/2}$, where p was the number of participants who were institutional/community recidivists, q was the number of participants who were institutional/community non-recidivists, and N was the total sample size.

The metrics used to estimate and interpret the magnitude of the relationships between each risk measure (e.g., HCR-20) or predictor category (e.g., 2nd generation measures) and misconduct or recidivism effect sizes were the mean r' value ($M_{r'}$) weighted by sample size (Z^+ , see Hedges & Olkin, 1985), along with its associated 95% confidence interval (CI_{Z^+}). The CI was used to reflect the degree to which there was agreement amongst study variables. If there was no overlap at all between the CI s for any two mean effect sizes, or the CI s just touched, then the effects would be interpreted as representing different population parameters. This criterion is equivalent to statistical significance of $p < .006$, as long as the sample size was ≥ 10 and the width of the CI s did not vary by more than a factor of two (Cumming & Finch, 2005). When CI s between any two comparisons overlapped by no more than one quarter of the average length of the two intervals, the mean effects were also viewed as representing two different population parameters and were statistically different at approximately $p \leq .05$ (see Cumming & Finch, 2005). Overlap in CI s that exceeded the above criterion meant that the mean effect sizes were likely produced from the sample population parameter, and therefore, not statistically different. A second use of CI s was to reflect the precision of effect size estimates, which was judged by noting the width of the CI (see Cumming & Finch, 2001; Gendreau, Goggin, & Smith, 2000; Schmidt, 1996). Narrower intervals indicate a more precise estimate of a population parameter than do wider intervals.

Effect Size Heterogeneity

The influence of outliers was determined using the Q statistic (Rosenthal, 1991). For each effect size, a q value was calculated using the formula: $(n - 3)(z_{r'} - Z^+)$; where n was the total sample size per effect size; $z_{r'}$ was the standardized r' value per effect size; and Z^+ was the sample-weighted $M_{r'}$ value for each predictor category. These q values were then summed for each predictor category, yielding Q , which is an estimate of the heterogeneity of the effect sizes within that category. To test its significance, the Q was evaluated using the critical value of χ^2 with $(k - 1)$ degrees of freedom. A significant Q statistic indicates that there is more variability than would be expected by chance. In such cases, outlying effect sizes were inspected and only eliminated if there was a logical reason for exclusion; for example, a coding error or unique study characteristic (e.g., restrictive sample).

Fail Safe Estimation

A fail safe estimate was employed to provide an index of how many additional effect sizes would be required to alter an obtained effect size estimate. An index of the number of effect sizes ($Z^+ = .00$) needed for a given risk measure of greater accuracy in the prediction of misconduct/recidivism to approach an effect size equal to one of lesser accuracy was calculated using the following formula: $[(k_B(Z_B^+ - Z_A^+))]/(Z_A^+ - Z_{B=0}^+)$, where $Z_{B=0}^+$ indicates a null effect for the more accurate risk measure

(see Gendreau, et al., 2002). As applied to the present meta-analysis, assume that the mean effect size for Measure A was .30 ($k = 50$) and .35 ($k = 40$) for Measure B. Using the above formula, an estimate of seven B predictions with an $Z^+ = 0$ would be necessary to negate Measure B's supremacy over A. That is, seven additional Measure B effect sizes, each with a magnitude of $Z^+ = .00$, would have to be located to conclude that the two measures were at predictive parity.

Results

Description of Database

The final dataset contained 88 studies reporting predictive validities for various risk measures with violent misconduct ($k = 76$) and violent recidivism ($k = 185$). Although the studies examined over 70 different risk measures in total, only those measures represented in 10 or more effect size estimates per outcome will be reported in order to emphasize the instruments for which the greatest amount of data was available. These instruments included the HCR-20 ($k = 11$ for misconduct; $k = 11$ for recidivism), LSI/LSI-R ($k = 19$ for recidivism), PCL/PCL-R ($k = 24$ for recidivism), SIR scale ($k = 17$ for recidivism); and the VRAG ($k = 14$ for recidivism). Some instruments with fewer than 10 effect sizes are reported where relevant, but their predictive validities should be interpreted cautiously given the low number of effect sizes.

Information used to complete the assessment instruments was predominantly collected via file extraction (52.2% of effect sizes), interview (11.2%), or a combination of the two methods (16.5%). Self-report assessment methods were used to gather data for 17.4% of effect sizes and were combined with an interview for less than 1% of effect sizes. A final 1.8% of effect sizes were derived from measures completed via behavioural observations by institutional staff. Risk assessment was entirely or predominantly based on static risk factors for 34.9% of effect sizes and on dynamic factors for 51.9% of effect sizes. Just over 8% of effect sizes were derived from measures using equal numbers of static and dynamic predictors. A final 4.9% came from studies where information was not sufficient to determine the nature of the predictors used. The vast majority of measures (85%) were relevant to corrections (i.e., rooted in a theory of criminal behaviour and/or created specifically for use as a criminal risk instrument). Just under 3% of effect sizes were based in clinical judgement (i.e., first generation assessment), 52.3% came from measures categorized as second generation actuarial measures. Third and fourth generation measures were represented in 42.3% and 2.5% of effect sizes, respectively.

In terms of study characteristics, 63.1% of effect sizes were recorded from studies published in books, journals, or government reports, 32% from thesis/doctoral dissertations, and 5% from unpublished data acquired directly from researchers (i.e., raw data, unpublished manuscripts, or conference paper/poster presentations). Canadian and U.S. studies produced 60.1% and 24.8% of effect sizes, respectively. Most effect sizes were reported in studies produced by academically affiliated authors (51.3%) from the discipline of psychology (85.4%). The effect sizes represented a combined total of 273,734 offenders (misconduct $N = 232,790$; recidivism $N = 40,944$). The majority (81.3% of effect sizes collapsed across both outcomes) came from predominantly male samples.

Samples representing general offender populations produced 63.9% of effect sizes, while the remainder were based on forensic (30.7%) and mixed (5%) samples. Institutional violence effect sizes were based on an equivalent percentage of general offender samples (50.7%) and forensic psychiatric samples (49.3%). In contrast, the majority of violent recidivism effect sizes were based on general offender samples (70.0%). A sample's risk level was defined either by the original study authors (6.3% of effect sizes) or by the principal coder for the current study (90.6% of effect sizes). Overall, most effect sizes were based on low or moderate risk samples (43.6% and 44.0% respectively). Only 7.5% of effect sizes came from high risk samples and just under 3% were based on mixed risk samples. For 2.1% of effect sizes, there was insufficient information from which to determine sample risk level. Predisposition for violence

among offenders could not be assessed with any degree of certainty because information about previous and current violent offences was not reported for 67.6% and 56.0% of effect sizes, respectively.

On average, the base rate for major violent institutional misconducts was 25.84% ($SD = 13.61\%$) and the mean violent recidivism base rate was 21.73% ($SD = 12.99$). Only 39.4% of institutional violence effect sizes were based on follow-up periods of greater than one year and most community released offenders were followed for periods of between 2 and 5 years (41.7%). The most common index of institutional misconduct was official prison records (74.7%), while re-arrest, re-conviction, or re-incarceration data were the most common violent recidivism indices (72.2% of effect sizes). In most studies (97.0%), violent recidivists were compared to an aggregate group of offenders (i.e., offenders who did not re-offend at all combined with those who may have non-violently re-offended). Thus, little predictive data was available using a pure “no recidivism at all” outcome criterion.

Risk Measures: Predictive Validities for Institutional Violence

Table 1 contains the Z^+ values and associated 95% CI s for risk measures and institutional violence outcomes. There was only one measure that was represented by more than 10 effect sizes (i.e., the HCR-20); however, to create consistency with the instruments reported for violent recidivism outcomes, preliminary data for congruent instruments are reported here despite a k of less than 10. The HCR-20 and LSI/LSI-R had the largest mean weighted effect sizes for predicting institutional violence ($Z^+ = .28$ and 24, respectively). The PCL:SV ($k = 7$) produced the third largest mean effect size ($Z^+ = .22$), while the PCL/PCL-R and VRAG produced the weakest associations with institutional violence ($Z^+ = .14$ and $Z^+ = .15$, respectively). However, the 95% CI s for each of the above risk measures overlapped considerably, suggesting that they were all sampling the same population parameter. Further, the width of their CI s and the small number of effect sizes foreshadow a lack of precision for each instruments' effect size estimate. As a result, interpretations based on these estimates should be viewed as tentative until more primary studies are conducted. Given that a minimum criterion of 10 effect sizes per instrument was set for calculation of fail safe analyses (to allow meaningful comparisons between the predictive validities of each instrument), these metrics were not calculated for institutional violence.

Risk Measures: Predictive Validities for Violent Recidivism

The Z^+ values with associated 95% CI s for predicting violent recidivism are displayed in the latter part of Table 1. The largest Z^+ value was recorded for the VRAG. There was CI overlap between this measure and the LSI-R and PCL-R, but its CI s did not overlap with the HCR-20 or SIR scale. Based on the widths of the CI s for each of these measures, the LSI-R, PCL-R, and SIR scale each generated slightly more precise point estimates than did the HCR-20 and the VRAG. Fail safe analyses indicated that six additional zero-effect VRAG effect sizes would be needed to reduce its predictive ability to that of the HCR-20 or SIR scale. Only another two null VRAG effect sizes would be needed for the VRAG to perform at par with the LSI-R or PCL-R.

In terms of notable measures with fewer than 10 effect sizes for violent recidivism (not reported in Table 1), the LS/CMI ($k = 3$, $N = 841$) yielded a strong magnitude of predictive validity ($Z^+ = .47$, $CI_{Z^+} = .40$ to $.54$), followed closely by the SAQ ($k = 8$, $N = 1094$, $Z^+ = .37$, $CI_{Z^+} = .31$ to $.43$). The CI s for these two measures only slightly overlapped and suggested that they may, in fact, be estimating distinct population parameters. Note, however, that any conclusions drawn about these two measures must be made in light of the fact that few effect sizes were available to test their predictive validity for violent recidivism. The

remaining notable measures were the PCL:SV ($k = 5, N = 641, Z^+ = .20, CI_{Z^+} = .12$ to $.28$) and measures comprised solely of criminal history items ($k = 9, N = 2230, Z^+ = .23, CI_{Z^+} = .19$ to $.27$). Based on the 3 effect sizes for the MMPI (using the Megargee Typology and the Prison Adjustment Scale), this instrument did not meaningfully predict violent recidivism ($Z^+ = .00$).

Table 1

Effect Size Comparisons of Risk Measures for the Prediction of Violent Institutional Misconduct^a and Recidivism^b

<i>Measure</i>	<i>k</i>	<i>N</i>	Z^+	CI_{z^+}	<i>Q</i>
Institutional Violence					
HCR-20	11	758	.28	.10 to .24	12.26
LSI/LSI-R	6	650	.24	.09 to .25	5.91
PCL/PCL-R	5	626	.14	.00 to .16	2.90
PCL:SV	7	504	.22	.07 to .25	5.59
VRAG	2	222	.15	-.08 to .18	1.54
Criminal History Indices	4	204132	.26	.24 to .25	12.83*
Violent Recidivism					
HCR-20	11	1395	.22	.17 to .27	6.68
LSI/LSI-R	19	4361	.28	.25 to .31	57.15*
PCL/PCL-R	24	4757	.27	.24 to .30	48.04*
SIR Scale	17	5618	.22	.19 to .25	32.54*
VRAG	14	2082	.32	.28 to .36	47.06*

* $p < .05$, indicates that there is more variability than would be expected by chance.

Note. k = effect sizes per risk measure; N = offenders per risk measure; Z^+ = r' value weighted by sample size; CI_{Z^+} = 95% confidence interval about Z^+ .

^a Although the total number of effect size estimates for risk measures with institutional violence was 76, there was only one category with $k > 10$. The other measures reported above are included to facilitate tentative comparisons of the predictive validity for those measures with misconduct and recidivism outcomes.

^b Although the total number of effect size estimates for risk measures with recidivism was 185, only those measures with more than 10 predictive validities were included in Table 1.

Comparison of Effect Sizes by Generation of Risk Instrument

Table 2 displays information relevant to comparing the mean effect size estimates for different generations of risk measures. The first and fourth generation measures had fewer than 10 effect sizes and, as such, were excluded from the table, but their preliminary data is described below. As shown in Table 2, the second generation instruments outperformed the third generation as predictors of institutional violence. This was due to the substantial weight given to three particularly large second generation studies with $ns > 10,000$ offenders. Fail safe calculations estimated that another 34 second generation effect sizes of zero would have to be added before its mean effect would be at par with that of the third generation measures in the prediction of institutional violence. The benefit of second versus third generation instruments was reversed when the outcome was violent recidivism; that is, third generation measures had a slight advantage over the second generation with no overlap of their CI s. According to the fail safe

index, another 23 null effect sizes for third generation studies would be needed to reduce its mean effect to that of the second generation instruments.

Table 2

Comparison of Risk Assessment Generations for the Prediction of Violent Institutional Misconduct^a and Recidivism^b

<i>Measure</i>	<i>k</i>	<i>N</i>	<i>Z</i> ⁺	<i>CI</i> _{Z⁺}	<i>Q</i>
Institutional Violence					
Second Generation	48	229397	.34	.33 to .35	410.12*
Third Generation	27	3349	.20	.17 to .23	26.94*
Violent Recidivism					
Second Generation	92	19874	.18	.17 to .19	328.13*
Third Generation	81	15233	.23	.21 to .25	247.38*

* $p < .05$, indicates that there is more variability than would be expected by chance.

Note. k = effect sizes per generation; N = offenders per generation; Z^+ = r' value weighted by sample size; CI_Z = 95% confidence interval about Z^+ .

^a Note that only 75 of 76 misconduct effect sizes were represented. One effect size, produced by a fourth generation measure, was not included in the table.

^b Note that only 173 of 185 recidivism effect sizes are represented. Seven effect sizes produced by a first generation measures and 5 effect sizes produced by fourth generation measure were not included in the table.

For the first and fourth generations (not referenced in Table 2), first generation methods produced a Z^+ of .18 ($k= 7$, $N = 1461$, $CI_{Z^+} = .13$ to .23) for violent recidivism, which was a higher effect size than expected based on the recidivism risk literature. A review of the data indicated that this estimate was based on only four studies, each of which used a different, but vaguely described, approach to render their clinical judgments of risk. In addition, four of these effects were generated from a single study. Of all the generations of instruments, fourth generation measures ($k = 5$, $N = 3759$) resulted in the largest predictive estimate ($Z^+ = .52$, $CI_{Z^+} = .49$ to .55) for violent recidivism. Notably, the fourth generation category shared no overlap with first, second, and third generation effect size estimates.

Comparisons Based on the Content of the Instrument: Static versus Dynamic

Table 3 summarizes the predictive accuracy of instruments that contained primarily static items, primarily dynamic items, and those with an equal combination of static and dynamic items. For institutional violence, static-based instruments generated the largest mean effect ($Z^+ = .32$) compared to the dynamic ($Z^+ = .21$) and the combined ($Z^+ = .23$) instruments, with no overlap of CI s between the primarily static and dynamic categories. The fail safe index noted that an additional 14 static effect sizes with an $r = 0$ would be needed to reduce the predictive magnitude of static-based instruments to the level of the dynamic-based instruments. Further, 10 additional studies would be necessary to reduce the predictive estimate of static instruments to that generated by the combination-based instruments. In terms of predicting violent recidivism, dynamic instruments had a slight predictive advantage over static instruments as evidenced by very little CI overlap between these factors (i.e., $p < .05$). The mean effect for dynamic instruments was marginally larger than combination instruments, with very slight overlap of the two CI s. Fail safe calculations indicated that another 13 dynamic effect sizes of zero would be needed

to reduce its predictive validity to that of static measures. An additional 24 nil effects sizes would be needed to reduce the predictive power of dynamic measures to that of combination measures.

Table 3

Comparison of Static and Dynamic-Based Instruments for Violent Institutional Misconduct^a and Recidivism^b

<i>Measure</i>	<i>k</i>	<i>N</i>	<i>Z</i> ⁺	<i>CI</i> _{Z⁺}	<i>Q</i>
Institutional Violence					
Static	26	226026	.32	.316 to .324	210.48*
Dynamic	37	5616	.21	.18 to .24	165.50*
Combination	12	1029	.23	.17 to .29	14.36
Violent Recidivism					
Static	64	13409	.22	.20 to .24	152.17*
Dynamic	96	21913	.25	.24 to .26	512.45*
Combination	13	1697	.20	.15 to .25	28.53*

* $p < .05$. The level of variability is greater than would be expected by chance.

Note. k = effect sizes per predictor domain; N = offenders per predictor domain; Z^+ = r' value weighted by sample size; CI_{Z^+} = 95% confidence interval about Z^+ .

^aNote that only 75 of the 76 institutional violence outcomes were represented. The nature of the predictors could not be determined for one of the effect sizes.

^bNote that only 173 of 185 recidivism effect sizes were represented. The nature of the predictors could not be determined for twelve of the effect sizes.

Comparisons Based on Measure Administration Method

Comparisons of the predictive validities between different administration methods are presented in Table 4. Beginning with institutional violence, the largest Z^+ value (.34) was attributed to the file extraction category (i.e., file review). Further, the CI associated with this mean effect shared no overlap with that of the self-report, interview, or combined file/interview categories. Fail safe calculation revealed that an additional 36 null file extraction effect sizes would have to be added to reduce its mean effect to that of the self-report category; a further 17 effect sizes of zero would be needed for parity with the combined file-interview method; and 46 nil effect sizes for equality with the interview method. Turning to recidivism outcomes, inspection of the second part of Table 4 showed that the file-interview method had the largest predictive validity ($Z^+ = .30$). The CI for this category only touched that of file extraction methods, and shared no overlap with the other two methods. To reduce the predictive accuracy of file-interview to that of file only, interview, or self-report methods, an additional 4, 47, or 41 nil file extraction effects, respectively, would be needed.

Table 4

Comparison of Administration Methods for the Prediction of Violent Institutional Misconduct^a and Recidivism^b

<i>Measure</i>	<i>k</i>	<i>N</i>	<i>Z</i> ⁺	<i>CI</i> _{Z⁺}	<i>Q</i>
Institutional Violence					
File extraction	32	223071	.34	.336 to .344	209.14*
Interview	6	635	.14	.06 to .22	2.52
Self-report	13	2505	.16	.12 to .20	21.63*
File + interview	13	1352	.22	.17 to .27	20.18
Violent Recidivism					
File extraction	97	24648	.26	.25 to .27	591.04*
Interview	21	2921	.11	.07 to .15	24.58
Self-report	29	5029	.12	.09 to .15	53.31*
File + interview	27	5741	.30	.27 to .33	100.13*

* $p < .05$. The level of variability is greater than would be expected by chance.

Note. k = effect sizes per method; N = offenders per method; Z^+ = r' value weighted by sample size; CI_{Z^+} = 95% confidence interval about Z^+ .^a Note that only 64 of the 76 institutional violence outcomes were represented. The nature of the administration method could not be determined for twelve of the effect sizes.

^b Note that only 174 of 185 recidivism effect sizes were represented. The nature of the administration method could not be determined for eleven of the effect sizes.

Comparisons Based on Instrument Relevance to Corrections

Another comparison of interest was the relevance of an instrument to corrections. Each effect size was coded as to whether the measure was derived from a criminological theory and/or whether it was created specifically for use as a risk instrument. For example, a measure like the LSI-R would have been coded as relevant to corrections because it was both derived from theories of criminality and created for use as a risk instrument. The VRAG also was coded as relevant because, although not created from theory, it was specifically created for risk evaluation.

Examples of non-relevant instruments were those designed to assess such constructs as levels of literacy, self-esteem, and psychiatric illness. Table 5 reports results for relevant versus non-relevant instruments in the prediction of both violent outcomes. Relevant instruments were better predictors for both institutional violence and recidivism, with no overlap in CI s with non-relevant instruments. Fail safe analyses indicated that, for institutional violence, an additional 19 effect sizes of zero for relevant instruments would be needed to reduce their predictive performance to that of non-relevant measures. For violent recidivism, as many as 415 additional null effects for relevant instruments would be needed to level the predictive validity between the two categories.

Table 5

Comparison of Relevant versus Non-Relevant Measures for the Prediction of Violent Institutional Misconduct and Recidivism^a

<i>Measure</i>	<i>k</i>	<i>N</i>	<i>Z</i> ⁺	<i>CI</i> _{<i>Z</i>⁺}	<i>Q</i>
Institutional Violence					
Relevant	63	214444	.35	.346 to .354	286.10*
Non-relevant	13	18346	.27	.26 to .28	144.25*
Violent Recidivism					
Relevant	153	33031	.26	.25 to .27	647.86*
Non-relevant	25	5835	.07	.04 to .10	56.93*

* $p < .05$. The level of variability is greater than would be expected by chance.

Note. k = effect sizes per predictor category; N = offenders per predictor category; Z^+ = r' value weighted by sample size; CI_{Z^+} = 95% confidence interval about Z^+ .

^a Note that only 178 of 185 recidivism effect sizes were represented. The relevance of the measures could not be determined for seven of the effect sizes.

Discussion

The prediction of violence has proven to be a challenging task for correctional and forensic professionals (Hanson, 2005). Fairly robust predictors of violence have been identified (e.g., Bonta et al., 1998) and instruments based on these factors have been developed to assist with risk prediction. Although professionals are presented with a range of tools to use, a challenge arises when trying to decide which of these instruments is most suitable for one's risk assessment purposes. In keeping with the risk-need-responsivity principles (Andrews & Bonta, 2003), it would be ideal to locate an instrument that maximizes the prediction of risk while still informing case management, rehabilitation planning, and progress in risk reduction. To assist with the decision-making process, the current meta-analysis synthesized the research focusing on the predictive validities of various instruments used to predict violence. From the pool of 88 studies that met our inclusionary criteria, a total of 185 effect sizes were produced for violent recidivism and 76 effect sizes were generated for violent institutional misconduct. Collapsed across instruments, the moderate ability to predict violent recidivism and violent institutional misconduct was consistent with estimates reported in other risk prediction meta-analyses (e.g., Gendreau et al., 1996; Gendreau, et al., 1997; Walters, 2006).

The following discussion should be considered with a mind to the limitations of the current meta-analysis. The first set of limitations related to the absence of necessary information within primary studies to code important variables as potential moderators. This is a familiar frustration to most meta-analysts. In the current study, 68% of effect sizes were based on studies in which there was insufficient information to code or define the level of violent history within a particular sample. In addition, none of the institutional violence studies provided details about their sample's pre-existing level of institutional violence. Thus, the extent of the omission of violent history data precluded an examination of this variable as a moderator of effect size. Furthermore, for 56% of effect sizes there was insufficient information about the nature of the sample's index offences (violent versus non-violent), which also prevented examination of the moderating effects of index offence severity on predictive validity. It was also interesting to note that 21 effect sizes were derived from studies that did not report the gender of the sample. When gender was noted, it was clear that most of the effect sizes were generated from male samples. Thus, generalization of the current results to female offenders, as well as to other poorly represented offender sub-groups (e.g., native offenders), is limited. One final, but important, methodological issue was that over 88% of effect sizes were generated from samples defined as low or moderate offending risk. Thus, it is difficult to generalize the current findings to high risk samples with any degree of certainty until additional data with this population has been accrued.

Actuarial, Structured, and Psychopathy Checklist Instruments

Violent Recidivism. Instruments based primarily on dynamic risk items generated the highest effect size magnitude for predicting violent recidivism ($Z^+ = .25$, see Table 3), and the *CI* for this category just touched the mean effect size *CI* for mostly static-based risk item measures ($Z^+ = .22$). Combined with the result of the fail-safe index, these findings suggest a small advantage for dynamic over static risk instruments when it comes to predicting violent re-offending. Similarly, third generation instruments produced a better estimate of violent recidivism risk than did second generation measures (see Table 2). Noting the limitations associated with only five effect sizes for fourth generation measures, this category of instruments produced the strongest predictive validities overall ($Z^+ = .52$). Interestingly, the predictive validity of unstructured clinical judgments of violent recidivism was higher than expected ($Z^+ = .18$).

given the negative view of this approach in the literature. However, it should be noted that this mean effect was based on seven effect sizes across only four different studies. Moreover, four of these effect sizes came from a single study (Rowe, 1995). It is possible that some of these judgments were made with a mind to relevant criminogenic risk factors given that these studies had been conducted at a time when significant information about risk prediction was available (primarily the 1990s). However, the means of formulating a clinical rating or risk judgment was not sufficiently mentioned by the authors of those four studies. Thus, it would be premature to conclude that unstructured clinical judgments of risk are valuable at this point. This is especially true in light of Hanson and Morton-Bourgon's (2007) recent meta-analysis, which demonstrates that the predictive validity of unstructured clinical predictions of violence risk among sex offenders is generally weak relative to that of actuarial prediction instruments.

In examining the mean effect size magnitudes for individual assessment tools with at least ten available effect sizes, it was clear that each was able to predict violent recidivism with at least moderate degree of success (see Table 1). The effect sizes ranged .22 for the HCR-20 and the SIR scale to .32 for the VRAG. The LSI/LSI-R, PCL-R, and SIR scales provided the most precise point estimates (i.e., produced the narrowest *CI*), but no one measure stood out as the most effective for predicting violent recidivism. The VRAG performed well, but the effect size *CI* overlapped with those of the LSI-R and the PCL-R. According to fail-safe indices, only two null VRAG effect sizes would be required to reduce its mean effect to that of the LSI-R and PCL-R. Thus, they are likely sampling the same population parameter. Further, the LSI-R was equivalent in its predictive validity to that of the PCL-R, and to a lesser degree with the HCR-20 and SIR scale. The VRAG performed better than the HCR-20 and the SIR scale. Thus, most of the measures reported in Table 1 appear to be similar in their predictive power. The exception being that the VRAG had a predictive advantage only over the HCR-20 and SIR scale.

The present results are congruent with past research, which has found that many of the commonly used risk instruments are moderately to highly inter-correlated (e.g., Glover et al., 2002). This implies that these measures share a significant portion of variance, although they do not completely overlap. The similarity between instruments was further reflected in a study by Kroner, Mills, and Reddon (2005), who randomly generated four hybrid risk measures based on the item content of the PCL-R, LSI-R, VRAG, and GSIR. When they tested each of these measures in terms of their ability to predict general recidivism, the hybrid instruments performed as well as each of the respective parent instruments. The current meta-analysis also updated an earlier finding of Gendreau et al. (2002), who reported a slight advantage of the LSI-R over the PCL-R in predicting violent recidivism. The inclusion of additional effect sizes published since Gendreau et al.'s data collection suggests that the PCL-R and the LSI-R are actually more comparable than not as predictors of violent re-offending.

In general, despite the justifiable concern about predicting future violence, and the ongoing debate as to which measure is best, there were still remarkably few effect sizes available to address these issues (i.e., the largest number was obtained for the PCL-R at $k = 24$). The present authors caution that there is likely little value in the generation of new risk measures at this point. The last thing the risk assessment field needs is to replicate the wasted efforts found in the psychiatric re-hospitalization prediction literature, in which 419 scales have been produced with only 3 reporting more than 10 predictive validity estimates (Smith, Gendreau, & Goggin, in press). Instead, research should focus on further validation of existing risk measures within different forensic contexts and offender sub-groups. Specifically, the majority of effect sizes used in the current analysis were based on a generic group of non-psychiatric offender samples, and the generalizability of the current findings to specialized offender groups requires additional study. Such information will likely better showcase an individual measure's strengths and weaknesses.

Researchers, in our opinion, should triple the number of effect sizes currently available before continuing the debate as to the supremacy of one measure over another in the prediction of violent recidivism.

Institutional Violence. In terms of institutional violence, second generation instruments had a predictive advantage over third generation measures ($Z^+ = .34$ vs. $.20$, respectively). Thus, instruments based on criminal history and other static variables were more informative than other measures when estimating the risk of institutional violence. Indeed, the current study found that instruments based primarily on static factors were better predictors of institutional violence ($Z^+ = .32$) than those based primarily on dynamic factors ($Z^+ = .21$), but both yielded more precise estimates of risk than instruments that equally combined static and dynamic factors. The Z^+ magnitude for the combined instrument did not meaningfully differ from the dynamic-based instruments. Both offender and forensic psychiatric samples were equally represented across the institutional violence effect sizes. As such, the nature of the sample is not necessarily contributing to the strength of static prediction tools for this outcome. It is possible that static factors are more valuable as risk items when assessing institutional violence because of the short-term duration of these assessments. Most of these effect sizes were based on studies with follow-up periods of less than 1 year, while the inclusion of dynamic risk factors may be more relevant to longer-term predictions of institutional violence (as they were for recidivism, which had longer follow-up periods).

Unlike the prediction of violent recidivism, there much more variability across the individual risk instruments in their ability to predict institutional violence. Criminal history indices ($Z^+ = .26$) were somewhat better predictors of institutional violence than any of the other measures, but this was a catch-all category for various measures relating to past criminality and its value is difficult to interpret as a result. In terms of standardized risk measures, the greatest number of effect sizes for the prediction of institutional violence were obtained for the HCR-20 ($Z^+ = .28$). However, it should be noted that despite its performance here, the HCR-20 has challenges related to its clinical application that are addressed later. In addition, the data for the HCR-20 was primarily based on forensic psychiatric samples, which limits its generalizability to institutional violence in non-psychiatric correctional facilities. In terms of other measures, the PCL:SV ($Z^+ = .22$) and LSI-R ($Z^+ = .24$) were also moderately predictive of institutional violence, while the poorly represented VRAG ($k = 2$) and PCL-R ($k = 5$) each recorded small associations with this outcome ($Z^+ = .15$ and $.14$, respectively). A few primary studies on the VRAG's predictive validity for institutional violence have come to light since the completion of our analyses (e.g., Nadeau, Nadeau, Smiley, & McHattie, 1999; McDermott, Edens, Quanbeck, Busse, & Scott, in press). The inclusion of these additional effect sizes in future meta-analyses may clarify the role of the VRAG in predicting institutional violence. Consistent with the current results regarding the PCL-R, Guy, Edens, Anthony, and Douglas (2005) found that the PCL-R produced a mean weighted effect size of $.17$ for physical aggression in the institution. In their analysis, the PCL-R was a better predictor of verbal aggression in the institution. Given the popularity of these latter two measures, it is hoped that additional results regarding their efficacy in institutions will be forthcoming. Therefore, with the exception of the HCR-20, LSI-R, and PCL:SV, considerable caution is warranted in the choice of instrument used to predict risk within an institutional setting until further prospective research has been conducted.

Other Relevant Findings. In general, the current data support the inclusion of self-report measures in the assessment of violence risk, but not as the sole means of prediction. Specifically, the mean effect size for the general category of self-report measures was small for both violent recidivism and institutional violence ($Z^+ = .12$ and $.16$, respectively), while the file review and file/interview approach produced the largest predictive validities for both outcomes. Among the self-report measures, one instrument that has received recent attention in the literature is the SAQ. The SAQ's prediction of institutional violence was

based on only one effect size, but suggested that it might have some utility in predicting that outcome ($r = .27$, Loza & Loza-Fanous, 2002). The mean effect size for the SAQ was very promising for violent recidivism ($Z^+ = .37$). The advantage of the SAQ is that it contains six scales that assess many of the empirically-identified risk-need factors for general and violent recidivism (Bonta et al., 1998; Gendreau et al., 1996), including antisocial attitudes, characteristics of antisocial personality disorder, early behaviour problems, past criminal behaviour, substance abuse, and antisocial associates. This instrument also contains a validity index and an anger subscale (Loza & Loza-Fanous, 2003). As found by Walters (2006), it is the content relevant self-report measures that are likely to yield more accurate estimates of violent risk and the SAQ fits this category. Nevertheless, given the SAQ's limited number of effect sizes, it requires additional primary studies to document its ability to prospectively predict violence and requires testing within different subgroups of offender populations (e.g., female offenders).

Similar to Walters' (2006) meta-analysis, measures assessing constructs unrelated or irrelevant to violent outcomes (e.g., anxiety) tended to perform poorly as predictors of violent recidivism in the current meta-analysis. Although the aggregate category of non-relevant measures did better at predicting institutional violence, they were still inferior to content-relevant measures in this category. Moreover, relatively little prospective data was available on their contribution to institutional violence. A notable finding within the current dataset was the little attention that the MMPI-2 has received as a predictor of future violent outcomes in recent prospective research. This is surprising because the MMPI was one of the most commonly used assessment instruments by psychologists working in correctional settings in the United States, albeit few of these professionals were actively engaged in risk assessment (Boothy & Clements, 2000). Although there are no specific statistics on the frequency of MMPI-2 use in Canadian violence risk assessments, professional practice indicates that it is certainly not a rare event. Only one, now dated, study was located on the predictive validity of the MMPI (Megargee Typology) as an index of future violence (Motiuk, 1991). This study found that it was a poor predictor of violent recidivism and only performed slightly better as a predictor of institutional violence. Thus, assessors must be cautious in the use of this instrument for informing decisions about risk given the lack of recent data regarding its predictive validity for violence. If one were to use the MMPI-2 in a violence risk assessment, it should be limited to understanding potential personality dynamics and mental health problems that may be relevant to responsivity concerns.

Recommendations to Guide Instrument Selection for the Assessment of Future Violence

An important practical issue for professional risk assessors and rehabilitators is the selection of the best instruments for their work with offenders (Bonta, 2002). Although the current analysis indicates that there was not much difference among the predictive validities of actuarial/structured instruments for violent re-offending, this does not mean that they would be equally informative for case planning and intervention strategizing when the goal is risk reduction. In light of the challenges with selecting appropriate instruments for violence risk assessment, several recommendations were developed. These recommendations were informed by the current meta-analysis and by professional practice parameters relevant to the selection of instruments for the purposes of violence risk assessment (see also Bonta, 2002; Quinsey et al., 1998). These recommendations stress the importance of considering the context and objective of the requested risk assessment, the content and structure of a particular risk instrument being considered for use, and the consideration of incorporating other measures to inform the scoring of standardized risk instruments and the formulation of risk reduction strategies.

Recommendation 1: Determine the Context and Purpose of the Risk Assessment. Instrument selection should depend on the objective and context of the setting in which the risk assessment is to be applied. For example, is the intent of the assessment to determine an offender's suitability for treatment, parole release, institutional placement, or security level? Or, is it to evaluate changes in the offender's risk and criminogenic needs over time? Further, is the assessment meant to inform case managers about an offender's behaviour within an institutional setting or in the community? Thus, it is important to select instruments that have been specifically designed for, and tested within, the context in which they are to be used. For instance, the promising VRS (Wong & Gordon, 2006) is meant to inform decisions about a high risk offender's candidacy for a violent offender treatment program; the Adult Internal Management System (AIMS; Quay, 1984) is designed to identify and manage offenders at risk for institutional violence while incarcerated; and the LSI-R (Andrews & Bonta, 1995) is intended to classify offenders in prison and community settings in a manner that better matches the intensity of supervision and rehabilitation services to the offender's risk-need level. Therefore, to identify relevant risk instruments, assessors must first clearly identify the context and purpose for which the offender's risk is being assessed.

A related issue in the selection of risk instruments concerns the determination of whether the objective of the assessment is to predict risk or inform and evaluate the effects of rehabilitation. If the end goal is pure prediction, with minimal interest in case planning beyond decisions about the level of supervision (e.g., decisions about security level), then almost any of the second generation instruments would be appropriate (e.g., the VRAG). However, as Heilbrun (1997) argued, risk assessment should be about more than mere prediction. Thus, if the objective is to predict risk *and* inform rehabilitation planning, then a third or fourth generation instrument (e.g., LSI-R, LS/CMI, HCR-20) is a more suitable candidate.

Recommendation 2: Consider the Content and Structure of Risk Instruments. From an assessment standpoint, it is the risk-related constructs that an instrument is supposed to measure that gives its risk score meaning and supplies a context in which to interpret that score (see Kroner et al., 2005). Thus, the assessor should choose an instrument that is based on solid theoretical constructs about risk and violence (the risk construct), while also giving consideration to the psychometric properties of the instrument in terms of its internal reliability, construct validity, and predictive validity. Of the instruments relevant to more than basic risk prediction, the HCR-20 and the LSI-R both were able to predict violence to a moderate degree in the current analysis and both have been developed to be theoretically and empirically-relevant to the assessment of risk.

Although the LSI-R was originally designed to assess general recidivism, and this *is* its strong suit (for a meta-analytic review see Gendreau et al., 2002), the current meta-analysis highlights its relevance to violence risk assessment. This is likely due to the substantial overlap in risk-need factors for violent and non-violent offenders (Bonta et al., 1998). In addition, both the LSI-R and the HCR-20 rely on multiple sources of information from which to derive their scoring of individual items (i.e., file reviews, collateral informants, and an interview with the offender), which is ideal for an informed assessment (Bonta, 2002). An additional advantage is that both may be sensitive to changes in risk because they contain some dynamic risk factors. Thus, either may be useful in the reassessment of risk over time.

Risk-need items within the HCR-20 and the LSI-R are combined to contribute to an overall estimate of risk, but it is at this point that a central difference emerges between the two measures. While the LSI-R uses a numerical estimate of risk (discussed below), the HCR-20 encourages the assessor to avoid relying on a numerical summation of item scores to derive the risk estimate. According to Webster et al. (1997b),

the presence of only a single risk item among the ten historical (e.g., substance abuse problems, psychopathy), five clinical (e.g., lack of insight, impulsivity), and five risk management (e.g., plans lack feasibility, non-compliance with remediation attempts) items included in their structured scheme could be sufficient to render a high risk judgment for a specific case. This would not be possible under a numerical system. Thus, assessors using the HCR-20 are advised to use it as a guide rather than an absolute actuarial tool. Nevertheless, Webster et al. (1997b) do acknowledge that the greater number of identified risk factors usually coincides with a greater risk of violence (see also Andrews & Bonta, 2003).

The HCR-20 has been criticized as being nothing more than a return to first generation clinical judgments of risk (Andrews et al., 2006) and, therefore, subject to the same errors and biases (Edens & Otto, 2001; Hilton et al., 2006). Research on this instrument, however, supports a less negative view. The current meta-analysis reported that the numerical risk scores generated by the HCR-20 (albeit for research purposes) were moderately predictive of future violence for prison and community settings. Past primary research reported that the HCR-20 structured judgments of risk (low, moderate, high) also showed acceptable levels of inter-rater reliability in retrospective designs, at least for psychiatric patients (de Vogel, de Ruiter, Hildebrand, Bos, & Van de Ven, 2004; Douglas, Ogloff, & Hart, 2003). This is not surprising given that the structured risk judgment in the HCR-20 is meant to be anchored in historical (i.e., temporally stable) risk factors. The structured judgment can then be adjusted on the basis of the dynamic clinical and risk management factors that may be present (Webster et al., 1997b).

Compared to the HCR-20, the LSI-R requires a more comprehensive analysis of risk-need factors. The assessor is required to rate 54 items, which are grouped into ten criminogenic risk-need areas, including criminal history, educational/employment issues, financial problems, family/marital concerns, accommodations, leisure/recreation, companions, alcohol/drug problems, emotional/personal problems, and the offender's attitude/orientation (Andrews & Bonta, 1995). Individual item ratings are summed to yield an overall LSI-R risk-need score, which is then converted into a percentile and compared to normative data supplied in the manual for both female and male offenders. Depending on where the offender's score falls in relation to this normative data, he or she is assigned a descriptive category ranging from low risk-needs to high risk-needs (Andrews & Bonta, 1995). The content of the LSI-R was modified in the LS/CMI by reducing the number of criminogenic risk-need domains to eight and the number of items to 43, as well as adding a specific "antisocial pattern" domain that captures aspects of psychopathic and antisocial personality characteristics (Andrews et al., 2004). There is also a space to note protective factors or offender strengths on the LS/CMI and a novel discharge summary form is also included. The new LS/CMI content and assessment forms are designed to more effectively integrate the risk assessment into the offender classification process, as well as into case management planning and the evaluation of offender progress.

A review of the items contained in the HCR-20 and LSI-R (and LS/CMI) indicate that they overlap on some item content (e.g., criminal history, antisocial/negative attitudes, employment problems, substance abuse). However, there are notable differences (e.g., HCR-20's focus on clinical issues). Another difference between the two measures concerns the role of the PCL-R. The HCR-20 includes an assessment of psychopathy using the PCL-R/SV, while the LSI-R does not. The PCL-based instruments are protected psychological tests and, as such, are not available to all correctional and forensic staff that may be required to complete risk assessments (e.g., parole officer). This limits the professional use of the HCR-20. It is possible to omit the PCL-R/SV item from consideration in the HCR-20 risk scheme, but the research is inconsistent regarding how this exclusion might affect the predictive validity of the scheme. Specifically, Grann et al. (2000) found that removal of the PCL item did not substantially impact on the

predictive validity of the Historical part of the HCR-20 in a retrospective follow-up of personality and mentally disordered offenders. In contrast, a retrospective study by Strand, Belfrage, Fransson, and Lavender (1999) found that, with the exception of the PCL-R item, the predictive validity of the Historical component of the HCR-20 suffered significantly in its ability to predict violent behaviour among forensic psychiatric patients. Thus, it is unclear whether the predictive power of the HCR-20 is reduced with the exclusion of the PCL-R/SV. An advantage of the LS/CMI is that its antisocial pattern domain contains some of the elements associated with psychopathic and antisocial personalities, but does not require administration of the PCL measures.

To conclude, the research to date suggests that the HCR-20 performs as well as the LSI-R for predicting violent behaviour in an aggregate forensic/general offender population. However, the effect sizes included in the present meta-analysis were based on the HCR-20's numerical scores and not on the structured clinical predictions advocated for use in its clinical application. Although assessors may be tempted to use the numerical scores for calculating risk estimates, it is inappropriate to interpret these scores without a proper normative group to which they can be compared. Thus, clinical use of the HCR-20 among offender populations should be restricted until more studies on the predictive validity of its structured judgments are conducted. Further, few prospective studies on the HCR-20 have been conducted with samples of non-mentally disordered offenders. As such, the applicability of the HCR-20 to non-mentally ill offenders is unclear. Finally, many of the HCR-20's validation studies only test the historical and risk management domains, rather than the clinical factors (Andrews & Bonta, 2003). Thus, additional research is required in general on its sub-components in both forensic and general offender samples.

As an alternative to the HCR-20, the LSI-R presents as a viable option for broad-based violence risk assessments that emphasize the assessment of criminogenic needs, rehabilitation planning, and the measurement of offender progress in risk reduction. Importantly, the point estimate for the LSI-R was rather concise for violent recidivism relative to the other measures. One value of the LSI-R, as noted by Gendreau et al. (2002), is its ability to be relevant to offenders who have criminal histories involving violent and non-violent crimes. Thus, there is a cost effective and time efficiency advantage to using an instrument that is capable of assessing general recidivism, as well as making a meaningful contribution to the prediction of violence. Unfortunately, the LSI-R and the LS/CMI only contain normative comparisons for general recidivism, not for violence risk. Thus, the inclusion of separate norms for violent recidivism would add value to the comprehensive nature of the LSI-R and the LS/CMI. Until these norms are generated, assessors should restrict their probabilistic risk estimates based on the LS/CMI and LSI-R to general recidivism. They can use their knowledge of the empirical research to indicate how this risk may also include violence within their qualitative descriptions of the risk estimate. The assessor should be sure to note the limitations for the violence risk estimate when communicating risk information to third parties.

Recommendation 3: Consider Content Relevant Measures as a Source of Information for Scoring Risk-Need Instruments. On a broader level than selecting the risk instrument, assessors should also decide on the relevance of information that will be used for the scoring of risk-need schemes. Consistent with the view that multi-method approaches (e.g., file review, interview, contact with collateral informants) increase the comprehensiveness of the information used to complete risk-need instruments (Andrews & Bonta, 2003; Andrews et al., 2006), the current analysis suggests that there are potential benefits to considering information gained from content-relevant self-report measures as part of the assessment process. This argument is also consistent with the findings of Walters (2006). Use of content-relevant self-report measures can provide information related to the assessment of risk that is difficult to assess by

other means, such as antisocial attitudes. In particular, the SAQ may add credibility to the assessor's subjective clinical impressions of an offender's cognitive distortions and antisocial belief systems. Validity indices on the SAQ help to minimize concerns with impression management on this scale. With further research, content relevant self-report measures, like the SAQ, could also potentially serve as selection tools for intervention programs and as pre-post measures in documenting changes in covert risk-needs domains. Only eight effect sizes were available for analysis, but it appears that the SAQ can contribute quite meaningfully to risk assessment. Despite its predictive power, caution is recommended in using the SAQ on its own to predict risk. As with any psychometric instrument, its score should be interpreted within the context of other known information about the individual in question. The current data clearly indicated the advantage of file and file/interview assessment methods over self-report and interviews for predicting risk. Thus, the SAQ can provide very relevant information for the assessment of risk and may inform scoring of criminogenic risk factors contained within other risk-need instruments.

Another psychological measure that might inform the scoring and interpretation of risk assessment is the PCL-R. This measure is not meant to be a risk instrument, despite its frequent use as such (Douglas, Vincent, & Edens, 2006). It was designed to measure a specific personality construct that is moderately related to violence, as demonstrated in the current and previous meta-analyses (e.g., Salekin, Rogers, & Sewell, 1996). This is why PCL-based scores have been included as items in such risk prediction instruments as the HCR-20 and the VRAG. Interestingly, the current data suggested that the PCL-R and PCL:SV performed differently in their ability to predict institutional violence. The PCL:SV, perhaps with its reduced emphasis on criminal history items and increased emphasis on psychopathic personality traits, appeared to be more appropriate for use in the assessment of institutional violence than the PCL-R. The PCL-R did provide a more precise effect size estimate than the PCL:SV for violent recidivism, but the two measures were similar in their predictive magnitudes for that criterion. They also performed similar to the standard risk instruments (e.g., VRAG, LSI-R, HCR-20) in predicting violent recidivism. Thus, the PCL-R and PCL:SV may not add incremental validity to the ability of actuarial measures to predict violence (Gendreau et al., 2002), but they can inform the case management strategies relevant to psychopathic offenders and their more challenging responsivity issues (e.g., egocentricity, manipulateness; Bonta, 2002; Douglas et al., 2006; Harris & Rice, 2006).

Further Issues for Future Research and Consideration

In addition to the need for the continued empirical validation of existing risk measures and the generation of violence risk norms for the LSI-R and LS/CMI, another area for future research is the identification of factors predictive of the nature and context of an offender's violent behaviour. Such research could identify acute/transitory risk factors relevant to determining the imminence of violence or assist with judgments about the likely occurrence of various forms of aggressive behaviour (e.g., reactive vs. instrumental; see Quinsey et al., 1998). The detailed aspects of violence risk, and the conditions under which violence is most likely to occur, are arguably more useful to case supervisors than a vague statement about the general estimate of violence risk. Recognizing the value of such qualitative information for violence risk management, these qualitative elements are often incorporated into descriptive formulations of risk in professional reports to stake holders. Unfortunately, despite a reasonable degree of proficiency at predicting the general likelihood of violent behaviour with the assistance of appropriate instruments, assessors have greater difficulty in accurately predicting the likelihood of various dimensions of violent behaviour (e.g., severity of aggression, likely imminence of the violent event, weapon use; Douglas et al., 2003). Thus, additional research is required on predictions

about the nature and quality of violence in order to validly inform the qualitative descriptions of risk provided to stakeholders.

Another often neglected aspect of violence risk assessment is the impact of protective factors, which may mediate or moderate the association between criminogenic factors and violent risk (see Rogers, 2000). This is partly due to the challenges in defining protective factors and in identifying their potential interactions with other variables, which together, lead to risk reduction. Hoge, Andrews, and Leschied (1996) conducted one of the few studies on this topic. Using a sample of young offenders on probation, Hoge and colleagues found that the presence of positive peer relationships, good achievement in school, positive response to authority, and an effective use of leisure time were associated with lower risk estimates on a youth version of the LSI-R. Youths with these protective factors also tended to be more compliant with the conditions of their probation and the effect of these factors were consistent for both low and high risk offenders. Thus, risk assessment instruments that incorporate protective factors would offer a more balanced evaluation of risk (see Rogers, 2000). Conceptualized as strengths, protective factors could serve as building blocks on which to facilitate changes that lead to risk reduction. The recognition of offender strengths is slowly gaining more interest as a component of fourth generation risk instruments (Andrews et al., 2006). Some structured prediction instruments (e.g., HCR-20) encourage consideration of protective factors in formulating final risk judgments as well. As of yet, there are no specific guidelines on how to go about this without over or underestimating risk. Furthermore, the adult and young offender risk literature is fairly limited in the ability to identify specific protective factors.

Conclusion

Based on the available research regarding its prospective predictive validity and its ability to inform case management and rehabilitation planning, the LSI-R (and likely the LS/CMI) proved to be a viable option for violence risk assessment. However, these instruments do not contain norms for violence risk in the same manner as they do for general recidivism. Thus, caution must be used in making probabilistic statements about the likelihood of violence when communicating the assessment results to others. Future research is needed to address this limitation. The HCR-20 also showed significant potential as a violence risk assessment tool, especially with forensic psychiatric populations. However, this tool requires additional research on the prospective validity of its structured prediction judgments and its numerical-based predictions with general offender populations. The PCL-R was a moderate predictor of violent recidivism, while the PCL:SV was a better predictor of institutional violence. The VRAG and SIR scales were successful in predicting violent recidivism, but they are not sensitive to changes in risk level that occur with time and/or rehabilitation efforts because of their static nature. In terms of self-report measures, the SAQ may be useful as a tool for inclusion within a risk assessment protocol to inform the scoring of criminogenic needs. Nonetheless, self-report and interviews should not be used on their own to assess risk. As a final statement, assessors need to be wary of instruments that have not been fully validated as meaningful predictors of *future* violence within the offender population in which they are planning to apply it. We also encourage researchers to continue building a database of prospective research concerning the prediction of violence across various offender settings and populations.

References

Entries marked with an asterisk (*) were included in the meta-analysis

- Andrews, D. A. (1989). Recidivism is predictable and can be influenced: Using risk assessments to reduce recidivism. *Forum on Corrections Research, 1*, 11-17.
- Andrews, D. A., & Bonta, J. (1995). *Level of Service Inventory-Revised*. Toronto, Ontario, Canada: Multi-Health Systems.
- Andrews, D. A., & Bonta, J. (2003). *The psychology of criminal conduct (3rd edition)*. Cincinnati, OH: Anderson.
- Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*, 19-52.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *The Level of Service/Case Management Inventory (LS/CMI)*. Toronto, Ontario, Canada: Multi-Health Systems.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency, 52*, 7-27.
- Andrews, D. A., & Wormith, J. S. (1984). *The Criminal Sentiments Scale*. Ottawa, Ontario, Canada: Ministry of Correctional Services of Canada.
- Baird, S. C. (1981). Classifying juveniles: Making the most of an important management tool. *Corrections Today, 47*, 32-38.
- Baird, S. C., Heinz, R. C., & Bemus, B. J. (1979). *The Wisconsin Case Classification/Staff development project: A two-year follow-up report*. Madison, WI: Wisconsin Bureau of Community Corrections.
- *Belfrage, H., Fransson, G., & Strand, S. (2000). Prediction of violence using the HCR-20: A prospective study in two maximum-security correctional institutions. *Journal of Forensic Psychiatry, 11*, 167-175.
- *Blanchette, K. (2005). *Field-test of a gender-informed security re-classification scale for female offenders*. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.
- *Blanchette, K., Verbrugge, P., & Wichmann, C. G. (2002). The Custody Rating Scale, initial security level placement, and women offenders. Research report #127. Ottawa, Ontario, Canada: Correctional Services of Canada.
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk-20: Professional guidelines for assessing risk for sexual violence*. Vancouver, British Columbia, Canada: The British Columbia Institute Against Family Violence.

- *Bonta, J. (1989). Native inmates: Institutional response, risk, and needs. *Canadian Journal of Criminology*, 31, 49-61.
- Bonta, J., (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice and Behavior*, 29, 355-379.
- *Bonta, J., Harman, W. G., Hann, R.G., & Cormier, R. B. (1996). The prediction of recidivism among federally sentenced offenders: A re-validation of the SIR scale. *Canadian Journal of Criminology*, 38, 61-79.
- Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin*, 123, 123-142.
- *Bonta, J., & Motiuk, L. L. (1986). Use of the Level of Supervision Inventory for assessing incarcerates. Paper presented at the annual convention of the American Psychological Association, Washington, D.C.
- *Bonta, J., & Yessine, A. K. (2005). The national flagging system: Identifying and responding to high-risk, violent offenders. Report No. 2005-04 to the Department of Public Safety and Emergency Preparedness Canada, Ottawa, Ontario, Canada.
- Boothy, J. L., & Clements, C. B. (2000). A national survey of correctional psychologists. *Criminal Justice and Behavior*, 27, 715-731.
- Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines, and training. *American Psychologist*, 51, 945-956.
- Capaldi, D. M., & Patterson, G. R. (1993). The violent adolescent male: Specialist or generalist? Paper presented at the 60th Biennial Meeting of the Society for Research in Child Development, New Orleans, LA, March 25-28.
- *Collie, R. M., & Polaschek, D. L. L. (2003). Sorting women's risk: New Zealand women prisoners' misconducts and internal security risk. *New Zealand Journal of Psychology*, 32, 101-109.
- *Cooke, D. (1996). Predicting offending in prison: The predictive validity of the prison behavior rating scales. *Legal and Criminological Psychology*, 1, 65-82.
- Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, 61, 532-574.
- Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, 60, 170-180.
- *Cunningham, M. D., & Sorensen, J. R. (2006). Actuarial models for assessing prison violence risk: Revisions and extensions of the Risk Assessment Scale for Prison (RASP). *Assessment*, 13, 253-265.

- *Cunningham, M. D., Sorensen, J. R., & Reidy, T. J. (2005). An actuarial model for assessment of prison violence risk among maximum security inmates. *Assessment, 12*, 40-49.
- *Daffern, M., Ogloff, J. R. P., Ferguson, M., & Thomson, L. (2005). Assessing risk for aggression in a forensic psychiatric hospital using the *Level of Service Inventory-Revised: Screening Version*. *International Journal of Forensic Mental Health, 4*, 201-206.
- *Dahle, K-P. (2006). Strengths and limitations of actuarial prediction of criminal reoffence in a German prison sample: A comparative study of LSI-R, HCR-20, and PCL-R. *International Journal of Law and Psychiatry, 29*, 431-442.
- de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B., & Van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health, 3*, 149-165.
- *Dernevik, M., Grann, M., & Johannson, S. (2002). Violent behaviour in forensic psychiatric patients: Risk assessment and different risk-management levels using the *HCR-20*. *Psychology, Crime, & Law, 8*, 93-111.
- *Dolan, M., & Khawaja, A. (2004). The HCR-20 and post-discharge outcome in male patients discharged from medium security in the UK. *Aggressive Behaviour, 30*, 469-483.
- Douglas, K. S., Hart, S. D., & Kropp, P. R. (2001). Validity of the Personality Assessment Inventory for forensic assessments. *International Journal of Offender Therapy and Comparative Criminology, 45*, 183-197.
- *Douglas, K. S., Ogloff, J. R. P., & Hart, S. D. (2003). Examination of a model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services, 54*, 1372-1379.
- Douglas, K. S. & Skeem, J. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law, 11*, 347-383.
- Douglas, K. S., Vincent, G. M., & Edens, J. F. (2006). Risk for criminal recidivism: The role of psychopathy. In C. J. Patrick (Ed.). *Handbook of psychopathy* (pp. 533-554). New York: Guilford.
- Douglas, K. S., Yeomans, N., and Boer, D.P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior, 32*, 479-510.
- Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent re-offending: A meta-analysis. *Canadian Journal of Criminology, 44*, 449-467.
- *Doyle, M., Dolan, M., & McGovern, J. (2002). The validity of North American risk assessment tools in predicting in-patient violent behaviour in England. *Legal and Criminological psychology, 7*, 141-154.
- Edens, J. F., & Otto, R. K. (2001). Release decision-making and planning. In J. B. Ashford, B. D., Sales, & W. Reid (Eds.), *Treating adult and juvenile offenders with special needs* (pp. 335-371). Washington, DC: American Psychological Association.

- *Edens, J. F., & Ruiz, M. A. (2006). On the validity of validity scales: The importance of defensive responding in the prediction of institutional misconduct. *Psychological Assessment, 18*, 220-224.
- French, S., & Gendreau, P. (2006). Reducing prison misconducts: What works! *Criminal Justice and Behavior, 33*, 185-218.
- *Fujii, D. E., Tokioka, A. B., Lichton, A. I., & Hishinuma, E. (2005). Ethnic differences in prediction of violence risk with the HCR-20 among psychiatric inpatients. *Psychiatric Services, 56*, 711-716.
- *Gagliardi, G. J., Lovell, D., Peterson, P. D., & Jemelka, R. (2004). Forecasting recidivism in mentally ill offenders released from prison. *Law and Human Behavior, 28*, 133-155.
- Gendreau, P., Goggin, C., French, S., & Smith, P. (2006). Practicing psychology in correctional settings. In I. B. Weiner & A. K. Hess (Eds.), *The handbook of forensic psychology* (3rd ed., pp. 722-750). Hoboken, NJ: Wiley & Sons.
- Gendreau, P., Goggin, C., & Law, M. A. (1997). Predicting prison misconducts. *Criminal Justice and Behavior, 24*, 414-431.
- Gendreau, P., Goggin, C., & Smith, P. (2000). Generating rational correctional policies: An introduction to advances in cumulating knowledge. *Corrections Management Quarterly, 4*, 52-60.
- Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the “unparalleled” measure of offender risk? A lesson in knowledge cumulation. *Criminal Justice and Behavior, 29*, 397-426.
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult recidivism: What works! *Criminology, 34*, 401-433.
- *Girard, L. (1999). *The Level of Service Inventory-Ontario Revision: Risk/need assessment and recidivism*. Unpublished doctoral dissertation, University of Ottawa, Ontario, Canada.
- *Glover, A. J. J., Nicholson, D. E., Hemmati, T., Berfeld, 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.
- Grann, M., Belfrage, H., & Tengström, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behavior, 27*, 97-114.
- *Grann, M., Långström, N., Tengström, A., & Kullgren, G. (1999). Psychopathy (PCL-R) predicts violent recidivism among criminal offenders with personality disorders in Sweden. *Law and Human Behavior, 23*, 205-218.
- *Gray, N. S., Hill, C., McGleish, A., Timmons, D., MacCulloch, M. J., & Snowden, R. J. (2003). Prediction of violence and self-harm in mentally disordered offenders: A prospective study of the efficacy of the HCR-20, PCL-R, and psychiatric symptomatology. *Journal of Consulting and Clinical Psychology, 71*, 443-451.

- *Gray, N. S., Snowden, R. J., MacCulloch, S., Phillips, H., Taylor, J., MacCulloch, T. J. (2004). Relative efficacy of criminological, clinical, and personality measures of future risk of offending in mentally disordered offenders: A comparative study of HCR-20, PCL:SV, and OGRS. *Journal of Consulting and Clinical Psychology, 72*, 523-530.
- *Grevatt, M., Thomas-Peter, B., & Hughes, G. (2004). Violence, mental disorder and risk assessment: Can structured clinical assessments predict the short-term risk of inpatient violence? *Journal of Forensic Psychiatry and Psychology, 15*, 278-292.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30.
- Guy, L. S., Edens, J. F., Anthony, C., & Douglas, K. S. (2005). Does psychopathy predict institutional misconduct among adults? A meta-analytic investigation. *Journal of Consulting and Clinical Psychology, 73*, 1056-1064.
- *Halsall, S., & Van Dine, S. (2002). Security reclassification risk instrument for males: Development and methodology. *Ohio Corrections Research Compendium, 1*, 95-96 Columbus, OH: Ohio Department of Rehabilitation and Correction.
- Hanson, R. K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence, 20*, 212-217.
- Hanson, R. K. & Morton-Bourgon, K. E. (2007). *The accuracy of recidivism risk assessment for sexual offenders: A meta-analysis*. Public Safety and Emergency Preparedness Canada.
- *Hanson, R. K., & Wallace-Capretta, S. (2000). *Predicting recidivism among male batterers* (User report 2000-06). Ottawa, Ontario, Canada: Solicitor General of Canada.
- Hare, R. D. (1991). *The Hare Psychopathy Checklist - Revised*. Toronto, Ontario, Canada: Multi-Health Systems.
- Hare, R. D. (2003). *The Hare Psychopathy Checklist – Revised (2nd edition)*. Toronto, Ontario, Canada: Multi-Health Systems.
- *Harer, M. D., & Langan, N. P. (2001). Gender differences in predictors of prison violence: Assessing the predictive validity of a risk classification system. *Crime and Delinquency, 47*, 513-536.
- Harris, G. T., & Rice, M. E. (2006). Treatment of psychopathy: A review of empirical findings. In C. J. Patrick (Eds.). *Handbook of Psychopathy* (pp. 555-572). New York: Guilford.
- *Harris, G. T., Rice, M. E., & Cormier, C. A. (1991). Psychopathy and violent recidivism. *Law and Human Behavior, 15*, 626-637.
- *Harris, G. T., Rice, M. E., & Cormier, C. A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law and Human Behavior, 26*, 377-394.

- *Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, *20*, 315-335.
- Hathaway, S. R., & McKinley, J. C. (1967). *Minnesota Multiphasic Personality Inventory manual (revised edition)*. New York: Psychological Corporation.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for analysis*. Toronto, Ontario, Canada: Academic Press.
- 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.
- *Heilbrun, K., Hart, S. D., Hare, R. D., Gustafson, D., Nunez, C., & White, A. J. (1998). Inpatient and post-discharge aggression in mentally disordered offenders: The role of psychopathy. *Journal of Interpersonal Violence*, *13*, 514-527.
- *Hemphill, J. F. (1998). *Psychopathy, criminal history, and recidivism*. Unpublished doctoral dissertation. University of British Columbia, Canada.
- Hemphill, J. F., Hare, R. D., & Wong, S. (1998). Psychopathy and recidivism: A review. *Legal and Criminological Psychology*, *3*, 139-170.
- *Hemphill, J. F., Templeman, R., Wong, S., & Hare, R. D. (1998). Psychopathy and crime: Recidivism and criminal careers. In D. J. Cooke, A. E. Forth, & R. D. Hare (Eds.), *Psychopathy: Theory, research and implications for society* (pp. 375-399). Dordrecht, Netherlands: Kluwer Academic.
- *Hildebrand, M., De Ruiter, C., & Nijman, H. (2004). PCL-R psychopathy predicts disruptive behavior among male offenders in a Dutch forensic psychiatric hospital. *Journal of Interpersonal Violence*, *19*, 13-29.
- Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on clinical versus actuarial prediction of violence. *The Counseling Psychologist*, *34*, 400-409.
- Hoffman, P. B. (1983). Screening for risk: A revised salient factor score (SFS 81). *Journal of Criminal Justice*, *11*, 539-547.
- Hoge, R. D., & Andrews, D. A. (1996). *Assessing the youthful offender: Issues and techniques*. New York: Plenum.
- Hoge, R. D., Andrews, D. A., & Leschied, A. W. (1996). An investigation of risk and protective factors in a sample of youthful offenders. *Journal of Child Psychology and Psychiatry*, *37*, 419-424.
- *Holland, T. R., Holt, N., Levi, M., & Beckett, G. E. (1983). Comparison and combination of clinical and statistical predictions of recidivism among adult offenders. *Journal of Applied Psychology*, *68*, 203-211.

- *Jayjohn, J., & Van Dine, S. (2002). Standard initial security classification risk instrument for males: Development and methodology. *Ohio Corrections Research Compendium, 1*, 93-94. Columbus, OH: Ohio Department of Rehabilitation and Correction.
- *Jemelka, R. P., Wiegand, G. A., Walker, E. A., & Trupin, E. W. (1992). Computerized offender assessment: Validation study. *Psychological Assessment, 4*, 138-144.
- *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*, 168-178.
- *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.
- Kroner, D. G., Mills, J. F., & Reddon, J. R. (2005). A coffee can, factor analysis, and prediction of antisocial behaviour: The structure of criminal risk. *International Journal of Law and Psychiatry, 28*, 360-374.
- Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1995). *Manual for the Spousal Assault Risk Assessment Guide (2nd edition)*. Vancouver, British Columbia, Canada: British Columbia Institute of Family Violence.
- *Law, M. A. (2004). *A longitudinal follow-up of federally sentenced women in the community: Assessing the predictive validity of the dynamic characteristics of the community intervention scale*. Unpublished doctoral thesis, Carleton University, Ottawa, Ontario, Canada.
- Ley, B. A. (1972). Quantitative aspects of psychological assessment: An introduction. London: Duckworth.
- Litwack, T. R., & Schlesinger, L. B. (1999). Dangerousness risk assessments: Research, legal, and clinical considerations. In A. K. Hess & I. B. Weiner (Eds.). *The handbook of forensic psychology* (pp. 171-217). New York: Wiley.
- *Lovell, D., Gagliardi, G., & Phipps, P. (2005). *Washington's dangerous mentally-ill offender law: Was community safety increased?* Report #05-03-10901. Olympia, WA: Washington State Institute for Public Policy.
- Loza, W., Dhaliwal, G. K., Kroner, D. G., & Loza-Fanous, A. (2000). Reliability and concurrent validity of the Self-Appraisal Questionnaire (SAQ): A tool for assessing violent and non-violent recidivism. *Criminal Justice and Behavior, 27*, 356-374.
- *Loza, W., & Green, K. (2003). The Self Appraisal Questionnaire: A self-report measure for predicting recidivism versus clinician-administered measures: A five-year follow-up study. *Journal of Interpersonal Violence, 18*, 781-797.
- *Loza, W., & Loza-Fanous, A. (1999b). *Predictive validity of the LSI: A tool for assessing violent and non-violent recidivism*. Unpublished manuscript.

- *Loza, W., & Loza-Fanous, A. (2000). Predictive validity of the Self-Appraisal Questionnaire (SAQ): A tool for assessing violent and nonviolent release failures. *Journal of Interpersonal Violence, 15*, 1183-1191.
- *Loza, W., & Loza-Fanous, A. (2001). The effectiveness of the Self-Appraisal Questionnaire in predicting offenders' post-release outcome: A comparison study. *Criminal Justice and Behavior, 28*, 105-121.
- *Loza, W., & Loza-Fanous, A. (2002). The effectiveness of the Self-Appraisal Questionnaire (SAQ) as an offenders' classification measure. *Journal of Interpersonal Violence, 17*, 3-14.
- *Loza, W., & Loza-Fanous, A. (2003). More evidence for the validity of the Self Appraisal Questionnaire for predicting violent and nonviolent recidivism: A five-year follow-up study. *Criminal Justice and Behavior, 30*, 709-721.
- *Loza, W., Villeneuve, D. B., & Loza-Fanous, A. (2002). Predictive validity of the Violence Risk Appraisal Guide: A tool for assessing violent offenders' recidivism. *International Journal of Law and Psychiatry, 25*, 85-92.
- *Luciani, F., Motiuk, L., & Nafekh, M. (1996). *An operational review of the Custody Rating Scale: Reliability, validity, and practical utility*. Research Report #R-47. Ottawa, Ontario, Canada: Correctional Services of Canada.
- *McHattie, L., Hills, A., Smiley, C., & MacKenzie, R. (1999). *Does the CATS predict recidivism as well as the PCL-R?* Abbotsford, British Columbia, Canada: Regional Health Centre.
- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111*, 361-365.
- Megargee, E., & Bohn, M. J., Jr. (1979). *Classifying criminal offenders: A new system based on the MMPI*. Beverly Hills, CA: Sage.
- Megargee, E. I., & Carbonell, J. L. (1995). Use of the MMPI-2 in correctional settings. In Y. S. Ben-Porath, J. R. Graham, G. C. N. Hall, R. D. Hirschman, & M. S. Zaragoza (Eds.), *Forensic applications of the MMPI-2* (pp. 127-159). Thousand Oaks, CA: Sage.
- *Mills, J. F., Jones, M. N., & Kroner, D. G. (2005). An examination of the generalizability of the LSI-R and VRAG probability bins. *Criminal Justice and Behavior, 32*, 565-585.
- *Mills, J. F., & Kroner, D. G. (2006). The effect of discordance among violence and general recidivism risk estimates on predictive accuracy. *Criminal Behaviour and Mental Health, 16*, 155-166.
- *Mills, J. F., Kroner, D. G., & Hemmati, T. (2004). The Measures of Criminal Attitudes and Associates (MCAA): The prediction of general and violent recidivism. *Criminal Justice and Behavior, 31*, 717-733.

- *Mills, J. F., Loza, W., & Kroner, D. G. (2003). Predictive validity despite social desirability: Evidence of the robustness of self-report among offenders. *Criminal Behaviour and Mental Health, 13*, 140-150.
- Monahan, J., & Steadman, H. (Eds.). (1994). *Violence and mental disorder: Developments in risk assessment*. Chicago: University of Chicago Press.
- Morey, L. C. (1991). *Personality Assessment Inventory: A professional manual*. Odessa, FL: Psychological Assessment Resources.
- Morey, L. C., & Quigley, B.D. (2002). The use of the Personality Assessment Inventory (PAI) in assessing offenders. *International Journal of Offender Therapy and Comparative Criminology, 46*, 333-349.
- *Motiuk, L. L. (1991). *Antecedents and consequences of prison adjustment: A systematic assessment and reassessment approach*. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.
- *Motiuk, M. S., Motiuk, L. L., & Bonta, J. (1992). A comparison between self-report and interview-based inventories in offender classification. *Criminal Justice and Behavior, 19*, 143-159.
- *Müller-Isberner, R., Sommer, J., Özokyay, K., & Freese, R. (1999, November). *Clinical use of the HCR-20 for predicting violence in a German Forensic Psychiatric Hospital*. Paper presented at the International Conference on Risk Assessment and Management: Implications for Prevention of Violence, Vancouver, British Columbia, Canada.
- Nadeau, J., Nadeau, B., Smiley, W. C., & McHattie, L. (1999, November). *The PCL-R, and VRAG as predictors of institutional behaviour*. Paper presented at Risk Assessment and Risk Management: Implications for the prevention of violence. Vancouver, British Columbia.
- *Nicholls, T. (2001). *Violence risk assessments with female NCRMD acquittees: Validity of the HCR-20 and PCL-SV*. Unpublished doctoral dissertation, Simon Fraser University, British Columbia, Canada.
- *Nicholls, T. L., Vincent, G. M., & Whitemore, K. E. (1999, November). *Assessing risk of inpatient violence in a sample of forensic psychiatric patients: Comparing the PCL:SV, HCR-20, and VRAG*. Paper presented at the International Conference on Risk Assessment and Management: Implications for Prevention of Violence, Vancouver, British Columbia, Canada.
- Nuffield, J. (1982). *Parole decision-making in Canada: Research towards decision guidelines*. Ottawa, Ontario, Canada: Ministry of Supply and Services Canada.
- *Nugent, P. M. (2000). *The use of detention legislation: Factors affecting detention decisions and recidivism among high-risk federal offenders in Ontario*. Unpublished doctoral dissertation, Queen's University, Kingston, Ontario, Canada.
- Osberg, T. M., & Poland, D. L. (2001). Validity of the MMPI-2 basic and Harris-Lingoes subscales in a forensic sample. *Journal of Clinical Psychology, 57*, 1369-1380.

- Phillips, H. K., Gray, N. S., MacCulloch, S. I., Taylor, J., Moore, S. C., Huckle, P., & MacCulloch, M. J. (2005). Risk assessment in offenders with mental disorders: Relative efficacy of personal demographic, criminal history, and clinical variables. *Journal of Interpersonal Violence, 20*, 833-847.
- Piquero, A. (2000). Frequency, specialization, and violence in offending careers. *Journal of Research in Crime and Delinquency, 37*, 392-418.
- *Polvi, N. H. (2001). *The prediction of violence in pre-trial forensic patients: The relative efficacy of statistical versus clinical predictions of dangerousness*. Unpublished doctoral dissertation, Simon Fraser University, British Columbia, Canada.
- Quay, H. (1984). *Managing adult inmates: Classification for housing and program assignments. Adult Internal Management System (AIMS) Classification manual*. College Park, MD: American Correctional Association.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and Managing Risk (1st edition)*. Washington, DC: American Psychological Association.
- *Raynor, P. (1998). *Implementing the Level of Service Inventory-Revised (LSI-R) in Britain: Report 3, risk and need assessment in the five pilot areas*. Centre for Criminal Justice and Criminology, Department of Social Policy and Applied Social Studies, University of Wales, Swansea, UK in association with The Cognitive Centre Foundation, Dinas Powys Vale of Glamorgan, UK.
- *Rettinger, L. J. (1998). *A recidivism follow-up study investigating risk and need within a sample of provincially sentenced women*. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.
- Rice, M. E. (1997). Violent offender research and implications for the criminal justice system. *American Psychologist, 52*, 414-423.
- *Rice, M. E., & Harris, G. T. (1992). A comparison of criminal recidivism among schizophrenic and nonschizophrenic offenders. *International Journal of Law and Psychiatry, 15*, 397-408.
- *Rice, M. E., & Harris, G. T. (1995). Psychopathy, schizophrenia, alcohol abuse, and violent recidivism. *International Journal of Law and Psychiatry, 18*, 333-342.
- Rogers, R. (2000). The uncritical acceptance of risk assessment in forensic practice. *Law and Human Behavior, 24*, 595-605.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Beverley Hills, CA: Sage.
- *Rowe, R. (1995). *The utilization of an interview-based classification instrument for parole board decision making in Ontario*. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.
- *Rowe, R. (1997). *Applying risk-needs assessments to parole decisions*. Manuscript in progress.

- Salekin, R. T., Rogers, R., & Sewell, K. W. (1996). A review and meta-analysis of the Psychopathy Checklist - Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice, 3*, 203-215.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training researchers. *Psychological Methods, 1*, 115-129.
- *Serin, R. T. (1996). Violent recidivism in criminal psychopaths. *Law and Human Behavior, 20*, 207-217.
- *Serin, R. C., & Brown, S. L. (1998). *Empirical guidelines for clinicians: Enhancing the practice of offender risk assessment*. Ottawa, Ontario, Canada: Research Branch, Correctional Service of Canada.
- Simourd, D. J. (1997). The Criminal Sentiments Scale-Modified and Pride in Delinquency scale: Psychometric properties and construct validity of two measures of criminal attitude. *Criminal Justice and Behavior, 24*, 52-70.
- *Simourd, D. J. (2004). Use of a dynamic risk/need assessment instruments among long-term incarcerated offenders. *Criminal Justice and Behavior, 31*, 306-323.
- *Simourd, D. J., & Van De Ven, J. (1999). Assessment of criminal attitudes: Criterion-related validity of the Criminal Sentiments Scale-Modified and the Pride in Delinquency Scale. *Criminal Justice and Behavior, 26*, 90-106.
- Smith, P., Gendreau, P., & Goggin, C. G. (in press). "What works" in predicting psychiatric hospitalization and relapse: The specific responsivity dimension of effective correctional treatment for mentally disordered offenders. In R. Ax & T. Faggan (Eds.), *Corrections, mental health, and social policy: International perspectives*. Springfield, IL: Charles C. Thomas & Ltd.
- *St. Amand, M. D. (2003). *The coping relapse model of recidivism and the prediction of failure after relapse*. Unpublished doctoral dissertation, Queen's University, Kingston, Ontario, Canada.
- *Stadtland, C., Kleindienst, N., Kröner, C., Eidt, M., & Nedopil, N. (2005). Psychopathic traits and risk of criminal recidivism in offenders with and without mental disorders. *International Journal of Forensic Mental Health, 4*, 89-97.
- *Strand, S., Belfrage, H., Fransson, G., & Lavender, S. (1999). Clinical and risk management factors in risk prediction of mentally disordered offenders – more important than historical data? A retrospective study of 40 mentally disordered offenders assessed with the HCR-20 Violence Risk Assessment Scheme. *Legal and Criminological Psychology, 4*, 67-76.
- *Stribling, A. L. (2003). *Exploring differences in clinical judgments of dangerousness of North Carolina psychiatric inpatients: Utility of the HCR-20 risk assessment scheme*. Unpublished doctoral dissertation, University of Tennessee.
- Swets, J. A. (1986). Indices of discrimination and diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99*, 100-117.

- *Tengström, A. (2001). Long-term predictive validity of historical factors in two risk assessment instruments in a group of violent offenders with schizophrenia. *Nordic Journal of Psychiatry*, *55*, 243-249.
- *Tengström, A., Grann, M., Långström, N., & Kullgren, G. (2000). Psychopathy (PCL – R) as a predictor of violent recidivism among criminal offenders with schizophrenia. *Law and Human Behavior*, *24*, 45-58.
- *Urbaniok, F., Noll, T., Grunewald, S., Steinbach, J., & Endrass, J. (2006). Prediction of violent and sexual offences: A replication study of the VRAG in Switzerland. *Journal of Forensic Psychiatry and Psychology*, *17*, 23-31.
- *Villeneuve, D. B., Oliver, N., & Loza, W. (2003). Cross-validation of the Self Appraisal Questionnaire with a maximum-security psychiatric population. *Journal of Interpersonal Violence*, *18*, 1325-1334.
- *Villeneuve, D. B., & Quinsey, V. L. (1995). Predictors of general and violent recidivism among mentally disordered inmates. *Criminal Justice and Behavior*, *22*, 397-410.
- Walters, G. D. (1995). The Psychological Inventory of Criminal Thinking Styles (PICTS). Part I: Reliability and preliminary validity. *Criminal Justice and Behavior*, *22*, 307-325.
- Walters, G. D. (1996). The Psychological Inventory of Criminal Thinking Styles (PICTS). Part III: Predictive validity. *International Journal of Offender Therapy and Comparative Criminology*, *40*, 105-112.
- Walters, G. D. (2006). Risk-appraisal versus self-report in the prediction of criminal justice outcomes. *Criminal Justice and Behavior*, *33*, 179-304.
- *Walters, G. D., Duncan, S. A., & Geyer, M. D. (2003). Predicting disciplinary adjustment in inmates undergoing forensic evaluation: A direct comparison of the PCL-R and PAI. *Journal of Forensic Psychiatry and Psychology*, *14*, 382-393.
- *Walters, G. D., & Mandell, W. (in press). Incremental validity of the Psychological Inventory of Criminal Thinking and Psychopathy Checklist: Screening Version in predicting disciplinary outcome. *Law and Human Behavior*.
- Walters, G. D., White, T. W., & Denney, D. (1991). The Lifestyle Criminality Screening Form: Preliminary data. *Criminal Justice and Behavior*, *18*, 406-418.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997a). Assessing risk of violence to others. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment, and treatment* (pp. 251-277). New York: Guilford.
- Webster, C. D., Douglas, K., Eaves, D., & Hart, D. (1997b). *HCR-20: Assessing risk for violence (Version 2)*. Vancouver, British Columbia, Canada: Mental Health, Law, & Policy Institute, Simon Fraser University.

- Webster, C. D., Harris, G. T., Rice, M. E., Cormier, C., & Quinsey, V. L. (1994). *The Violence Prediction Scheme: Assessing dangerousness in high risk men*. Toronto, Ontario, Canada: University of Toronto, Centre of Criminology.
- *Wintrup, A. M. (1996). *Assessing risk of violence in mentally disordered offenders with the HCR-20*. Unpublished doctoral dissertation. Simon Fraser University, Burnaby, British Columbia, Canada.
- *Wong, S. C. P., & Gordon, A. (2001). The Violence Risk Scale. *Forensic Update*, 67, 19-23.
- *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 the Law*, 12, 279-309.
- *Wormith, J. S., Olver, M. E., Stevenson, H. E., & Girard, L. (2006). *The long term prediction of offender recidivism using diagnostic, personality, and risk/need approaches to offender assessment*. Manuscript in preparation.
- Yeaton, W.H., & Wortman, P.M. (1993). On the reliability of meta-analytic reviews: The role of intercoder agreement. *Evaluation Review*, 17, 292-309.

Appendix A

Studies Included in the Meta-Analysis and their Respective Effect Size Estimates for the Prediction of Institutional Violence (iv) and Recidivism (recid)

No.	Study	Measure	N _{iv}	N _{recid}	r _{iv}	r _{recid}
1	Belfrage et al. (2000)	Historical-Clinical-Risk Scale-20	41	-	.44	-
		Psychopathy Checklist: Screening Version	41	-	.33	-
2	Blanchette (2005)	Security Reclassification Scale for Women	400	-	.27	-
		Offender Security Level Rating	400	-	.24	-
3	Blanchette et al. (2002)	CRS: Institutional Adjustment	61	-	.39	-
		CRS: Institutional Adjustment	230	-	.19	-
		Custody Rating Scale: Security Risk	61	-	.01	-
		Custody Rating Scale: Security Risk	230	-	.18	-
4	Bonta (1989)	LSI/LSI-R	49	-	.16	-
		LSI/LSI-R	71	-	.25	-
5	Bonta et al. (1996)	SIR Scale	-	3267	-	.15
6	Bonta & Yessine (2005)	SIR Scale	-	159	-	.43
		VRAG	-	48	-	.31
		VRAG-Proxy	-	207	-	.33
		LSI:SV	-	235	-	.27
		STATIC-99	-	154	-	.14
7	Bonta & Motiuk (1986)	LSI/LSI-R	119	-	.36	-
8	Collie & Polaschek (2003)	Australian Security Classification Instrument	889	-	.25	-
9	Cooke (1996)	PBRs: Anti-authority	220	-	.15	-
		PBRs: Anti-authority	220	-	.08	-
		PBRs: Anti-authority	220	-	.10	-
10	Cunningham & Sorensen (2006)	PBRs: Anxious-depressed	13341	-	.22	-
11	Cunningham et al. (2005)	PBRs: Anxious-depressed	2505	-	.24	-
12	Daffern et al. (2005)	LSI: Screening Version	232	-	.15	-
13	Dahle (2006)	LSI/LSI-R	-	307	-	.23
		Historical-Clinical-Risk Scale-20	-	307	-	.31
		PCL/PCL-R	-	307	-	.32
14	Dernevik et al. (2002)	Historical-Clinical-Risk Scale-20	54	-	.46	-
		Psychopathy Checklist: Screening Version	54	-	.17	-
15	Dolan & Khawaja (2004)	Historical-Clinical-Risk Scale-20	-	70	-	.44
16	Douglas et al. (2003)	Historical-Clinical-Risk Scale-20	-	100	-	.26
17	Doyle et al (2002)	Psychopathy Checklist: Screening Version	87	-	.38	-
		VRAG using PCL:SV Psychopathy	87	-	.22	-
		Historical-Scale-10	87	-	.25	-
18	Fujji et al. (2005)	Historical-Clinical-Risk Scale-20	41	-	.09	-
		Historical-Clinical-Risk Scale-20	38	-	.18	-
		Historical-Clinical-Risk Scale-20	29	-	.35	-
19	Gagliardi et al. (2004)	Various criminal history items	-	333	-	.42
20	Girard (1999)	LSI/LSI-R	-	630	-	.28
		LSI/LSI-R	-	36	-	.24
21	Glover et al. (2002)	SIR Scale	-	106	-	.34
		VRAG	-	106	-	.32
		VRAG-Child Adolescent Taxon	-	106	-	.29
		Conduct disorder diagnosis	-	106	-	.28
		Violent SIR Scale	-	106	-	.27
		Psychiatric Referral Screening Form	-	106	-	.25
		Child Adolescent Taxon	-	106	-	.18

No.	Study	Measure	N _{iv}	N _{recid}	r _{iv}	r _{recid}
		PCL/PCL-R	-	106	-	.10
		DSM III/IV Antisocial Personality Disorder	-	106	-	.03
22	Grann et al. (1999)	PCL/PCL-R	-	352	-	.36
23	Gray et al. (2003)	Historical-Clinical-Scale-15	34	-	.53	-
		PCL/PCL-R	34	-	.35	-
		Beck Hopelessness Scale	34	-	.18	-
		Brief Psychiatric Rating Scale	34	-	.61	-
24	Gray et al. (2004)	Psychopathy Checklist: Screening Version	-	316	-	.11
		Historical-Clinical-Risk Scale-20	-	316	-	.08
		Offender Group Reconviction Scale	-	316	-	.08
25	Grevatt et al. (2004)	Historical-Clinical-Scale-15	44	-	.10	-
		Violence Risk Scale	44	-	.05	-
26	Hanson & Wallace-Capretta (2000)	LSI/LSI-R	-	275	-	.32
27	Harer & Langen (2001)	Various criminal history items	177767	-	.18	-
		Various criminal history items	24675	-	.15	-
28	Harris et al. (1991)	PCL/PCL-R	-	169	-	.42
		LSI/LSI-R	-	169	-	.24
29	Harris et al. (2002)	VRAG	-	133	-	.48
		Clinical assessment	-	383	-	.17
30	Harris et al. (1993)	PCL/PCL-R	-	618	-	.34
		LSI/LSI-R	-	618	-	.23
		Schizophrenia diagnosis	-	618	-	-.17
		Personality disorder diagnosis	-	618	-	.26
		VRAG	-	618	-	.42
		Various criminal history items	-	618	-	.36
31	Heilbrun et al. (1998)	PCL/PCL-R	218	181	.14	.16
32	Hemphill (1991)	PCL/PCL-R	-	106	-	.06
		SIR Scale	-	106	-	.00
		Salient Factor Score	-	106	-	.10
33	Hemphill et al. (1998)	PCL/PCL-R	-	274	-	.20
34	Hildebrand et al. (2004)	PCL/PCL-R	92	-	.03	-
35	Holland et al. (1983)	Clinical assessment	-	198	-	.37
		Salient Factor Score	-	198	-	.19
36	Jemelka et al. (1992)	Personality disorder diagnosis	500	-	.65	-
37	Kroner & Loza (2001)	Self Appraisal Questionnaire	-	78	-	.30
		PCL/PCL-R	-	78	-	.21
		SIR Scale	-	78	-	.30
		VRAG	-	78	-	.14
38	Kroner & Mills (2001)	PCL/PCL-R	97	87	.14	.12
		LSI/LSI-R	97	87	.20	.19
		Historical-Clinical-Risk Scale-20	97	87	.11	.16
		VRAG	97	87	.26	.11
		Lifestyle Criminality Screening Form	97	87	.13	.12
39	Law (2004)	Community Intervention Scale	-	497	-	.11
40	Lovell et al. (2005)	LSI/LSI-R	-	100	-	.21
		Various criminal history items	-	100	-	.37
41	Loza & Green (2003)	SIR Scale	-	91	-	.35
		LSI/LSI-R	-	91	-	.22
		PCL/PCL-R	-	91	-	.22
		Self Appraisal Questionnaire	-	91	-	.30
		VRAG	-	91	-	.19
42	Loza & Loza-Fanous (1999)	LSI/LSI-R	-	140	-	.11
43	Loza & Loza-Fanous (2000)	Self Appraisal Questionnaire	-	153	-	.32
44	Loza & Loza-Fanous (2001)	SIR Scale	-	68	-	.32

No.	Study	Measure	N _{iv}	N _{recid}	r _{iv}	r _{recid}
		LSI/LSI-R	-	68	-	.23
		PCL/PCL-R	-	68	-	.19
		Self Appraisal Questionnaire	-	68	-	.32
		VRAG	-	68	-	.22
45	Loza & Loza-Fanous (2002)	Self Appraisal Questionnaire	303	-	.27	-
46	Loza & Loza-Fanous (2003)	Self Appraisal Questionnaire		305	-	.34
47	Loza et al. (2002)	VRAG		124	-	.05
48	Luciani et al. (1996)	Custody Rating Scale: Total Score	2187	-	.19	-
49	McHattie et al. (1999)	Child Adolescent Taxon	-	42	-	.06
		PCL/PCL-R	-	42	-	.27
50	Mills et al (2005)	VRAG	-	209	-	.26
51	Mills & Kroner (2006)	PCL/PCL-R	-	209	-	.18
		LSI/LSI-R	-	209	-	.26
		SIR Scale	-	209	-	.30
52	Mills et al. (2004)	MCAA-Part A	-	144	-	.18
		MCAA-Part B	-	144	-	.30
		SIR Scale	-	144	-	.38
53	Mills et al. (2003)	Self Appraisal Questionnaire	-	77	-	.37
		SIR Scale	-	77	-	.29
		BIDR-Impression Management	-	77	-	.38
		BIDR-Self -deception	-	77	-	.24
54	Motiuk (1991)	MMPI/MMPI-II	215	215	.14	.02
		MMPI/MMPI-II	215	215	.14	.06
		MMPI/MMPI II	215	215	.04	.06
		Wisconsin Assessment of Client Risk	215	215	.10	.13
		Illinois Initial Risk Scale	215	215	.19	.08
		LSI/LSI-R	215	215	.18	.17
		Salient Factor Score	215	215	.18	.16
		SIR Scale	215	215	.08	.08
		Oregon Parole Prognosis	215	215	.16	.12
		Pennsylvania Parole Prognosis	215	215	.03	.09
55	Motiuk et al (1992)	LS:SRI	99	-	.28	-
		LSI/LSI-R	99	-	.32	-
56	Müller-Isberner et al. (1999)	Historical-Clinical-Risk Scale-20	220	-	.23	-
57	Nicholls (2001)	Historical-Clinical-Risk Scale-20	31	39	.26	.26
		Historical-Clinical-Risk Scale-20	30	39	.43	.15
		Psychopathy Checklist: Screening Version	31	39	.27	.32
		Psychopathy Checklist: Screening Version	30	39	.32	.37
58	Nicholls et al. (1999)	Historical-Clinical-Risk Scale-20	125	-	.31	-
		VRAG	125	-	.08	-
		Psychopathy Checklist: Screening Version	125	-	.14	-
59	Nugent (2000)	Historical-Clinical-Risk Scale-20	-	120	-	.18
		SIR Scale	-	120	-	.23
		LSI/LSI-R	-	120	-	.21
		VRAG	-	120	-	.23
		PCL/PCL-R	-	120	-	.17
		Child Adolescent Taxon	-	120	-	.09
60	Jayjohn & Van Dine (2002)	Various criminal history items	800	-	.35	-
61	Halsall & Van Dine (2002)	Various criminal history items	800	-	.39	-
62	Polvi (1999)	Clinical Assessment	-	215	-	.12
		Dangerous Behavior Rating Scale	-	215	-	.14
		Historical-Clinical-Risk Scale-20	-	215	-	.09
		Psychopathy Checklist: Screening Version	-	215	-	.25
		VRAG	-	215	-	.31

No.	Study	Measure	N _{iv}	N _{recid}	r _{iv}	r _{recid}
63	St. Amand (2002)	SIR Scale	-	157	-	.10
		PCL/PCL-R	-	164	-	.23
		Various criminal history items	-	159	-	.09
		Various criminal history items	-	159	-	.27
		Various criminal history items	-	159	-	.14
		Social Problem-Solving Interview	-	164	-	.27
		Access to criminal resources items	-	121	-	.16
		CSLQ	-	130	-	.06
		Violent Beliefs Inventory	-	164	-	.19
		SIR Scale	-	234	-	.16
		Child Adolescent Taxon	-	233	-	.06
		PCL/PCL-R	-	233	-	.05
		Various criminal history items	-	234	-	.00
		Various criminal history items	-	234	-	.17
		Various criminal history items	-	234	-	.20
		Problem Survey Checklist	-	234	-	.01
		Impulsiveness Questionnaire	-	233	-	.02
		Positive Affect /Negative Affect	-	224	-	.02
		Positive Affect/ Negative Affect	-	224	-	.09
		Anxiety items	-	229	-	.09
		Perceived Problem Index	-	232	-	.07
		Coping items	-	231	-	.02
		Coping items	-	231	-	.04
		Social Problem Solving Interview	-	231	-	.04
		Criminal Insensitivity/Irresponsibility Scale	-	224	-	.10
		Access to criminal resources items	-	217	-	.01
		CSLQ	-	227	-	.00
		Social Support Scheme	-	231	-	.09
		Time Use Questionnaire	-	220	-	.07
		Violent Beliefs Inventory	-	233	-	.01
		BIDR-Impression Management	-	207	-	.14
		BIDR-Self-deception	-	164	-	.16
		64	Serin (1996)	PCL/PCL-R	-	81
SIR Scale	-			81	-	.08
Salient Factor Score	-			81	-	.15
Base Expectancy Score	-			81	-	.21
65	Serin & Brown (1998)	PCL/PCL-R	-	263	-	.31
66	Simourd (2004)	LSI/LSI-R	-	129	-	.26
67	Simourd & Van De Ven (1999)	CSS/CSS-M	-	87	-	.17
		CSS/CSS-M	-	54	-	.07
		Pride in Delinquency Scale	-	87	-	.03
		Pride In Delinquency Scale	-	54	-	.04
68	Stadtland et al. (2005)	PCL/PCL-R	-	258	-	.25
69	Strand et al. (1999)	Historical-Clinical-Risk Scale-20	-	40	-	.51
		Psychopathy Checklist: Screening Version	-	40	-	.35
70	Stribling (2003)	Historical-Clinical-Risk Scale-20	52	-	.51	-
71	Tengström (2001)	Historical-Scale-10	-	106	-	.41
		VRAG	-	106	-	.29
72	Tengström et al. (2000)	PCL/PCL-R	-	202	-	.36
73	Urbaniok et al. (2006)	VRAG	-	79	-	.34
74	Villeneuve et al. (2003)	Self Appraisal Questionnaire	-	49	-	.28
		Self Appraisal Questionnaire	-	273	-	.33
75	Villeneuve & Quinsey (1995)	SIR Scale	-	117	-	.25
		Violence Recidivism Scale (VRISK)	-	117	-	.43

No.	Study	Measure	N _{iv}	N _{recid}	r _{iv}	r _{recid}
76	Walters & Mandell (in press)	Psychopathy Checklist: Screening Version	136	-	.14	-
		PICTS	136	-	.19	-
		PICTS: P Scale	136	-	.24	-
		PICTS: R Scale	136	-	.16	-
77	Wintrup (1996)	PCL/PCL-R	-	70	-	.20
		Historical-Clinical-Risk Scale-20	-	70	-	.20
78	Wong & Gordon (2001)	Violence Risk Scale (VRS)	-	2000	-	.46
79	Wong & Gordon (2006)	Violence Risk Scale (VRS)	-	918	-	.40
80	Wormith et al. (2006)	LS/CMI	-	60	-	.31
		PCL/PCL-R	-	60	-	.28
		DSM III/IV Antisocial Personality Disorder	-	60	-	.35
81	Raynor (1998)	LSI/LSI-R	-	147	-	.26
82	Rettinger (1998)	LSI/LSI-R	-	441	-	.44
		LS/CMI	-	441	-	.44
83	Rice & Harris (1992)	Schizophrenia diagnosis	-	190	-	.10
		LSI/LSI-R	-	618	-	.11
		Personality disorder diagnosis	-	618	-	.11
84	Rice & Harris (1995)	PCL/PCL-R	-	190	-	.27
		Substance abuse items	-	190	-	.10
85	Rowe (1995)	LSI/LSI-R	-	389	-	.34
		SIR Scale	-	389	-	.24
		Salient Factor Score	-	289	-	.14
		Clinical assessment	-	262	-	.15
		Clinical assessment	-	145	-	.11
		Clinical assessment	-	112	-	.11
		Clinical assessment	-	146	-	.13
86	Rowe (1997)	LS/CMI	-	340	-	.30
87	Walters et al. (2003)	PCL/PCL-R	185	-	.11	-
		PAI-Aggression subscale	149	-	.17	-
		PAI-Antisocial subscale	149	-	.12	-
88	Edens & Ruiz (2006)	PAI-Antisocial subscale	349	-	.15	-
		PAI-Positive Impression Management	349	-	.02	-

Note. BIDR = Balanced Inventory of Desirable Responding; CRS = Custody Rating Scale; CSLQ = Criminal Socialisation and Lifestyle Questionnaire; CSS/CSS-M = Criminal Sentiments Scale/ Criminal Sentiments Scale-modified; LS/CMI = Level of Service/Case Management Inventory; LSI/LSI-R = Level of Supervision or Level of Service Inventory-Revised; MCAA = Measures of Criminal Attitudes and Associates; PAI = Personality Assessment Inventory; PBRS = Prison Behavior Rating Scale; PICTS = Psychological Inventory of Criminal Thinking Styles; PCL/PCL-R = Psychopathy Checklist/Psychopathy Checklist-Revised; SIR Scale = Statistical Information on Recidivism Scale; VRAG = Violence Risk Assessment Guide;