Performance Indicators of the Structured Professional Judgment Approach for Assessing Risk for Violence to Others: A Meta-Analytic Survey

by
Laura S. Guy

B.A., McGill University, 1998
M.A., Sam Houston State University, 2002

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the

Department of Psychology

Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY

Fall 2008

APPROVAL

Name: **Laura S. Guy**

Degree: **Doctor of Philosophy**

Title of Thesis: **Performance Indicators of the Structured Professional Judgment Approach for Assessing Risk for Violence to Others: A Meta-Analytic Survey**

Examining Committee:

Chair: **Dr. Wendy Thornton**
Professor of Psychology

Dr. Kevin S. Douglas
Senior Supervisor
Associate Professor

Dr. Stephen D. Hart
Supervisor
Professor

Dr. John F. Edens
Supervisor
Associate Professor of Psychology
Texas A&M University

Dr. Patrick Lussier
Internal Examiner
Assistant Professor

Dr. Alec Buchanan
External Examiner
Associate Professor
Yale University

Date Defended/Approved: October 29, 2008

ABSTRACT

Tremendous advancements have been realized during the past several decades in the science and practice of the field now known as violence risk assessment. Whereas in the 1970s and 1980s estimates (dichotomous predictions) of individuals' potential to act violently tended to be based on unstructured clinical judgment, new technologies, or risk assessment tools, were developed during subsequent decades to assist professionals conducting such assessments. Initial technologies available were actuarial in nature; these efforts were followed by clinically based tools developed according to the structured professional judgment (SPJ) model with the intent of overcoming the perceived limitations of the actuarial approach. Throughout the field's metamorphosis, a steadfast theme has been impassioned commentary regarding the relative merit of actuarial and clinical approaches. Although much research has examined specific SPJ tools, to date, a comprehensive evaluation of the SPJ decision making model has not been conducted. This dissertation applied meta-analytic techniques to examine the predictive validity of the SPJ model using 113 disseminations. Results supported the utility of the SPJ model (especially when summary risk ratings were used) and indicated no distinct superiority for either the actuarial or SPJ model among the 44 samples in which direct comparisons of both approaches were made. It is concluded that both types of technologies perform at comparably good levels of predictive accuracy, but that additional factors are relevant when selecting an approach for clinical practice. Implications for practice and research are discussed.

Keywords: Structured professional judgment; violence risk assessment; violence risk management; hcr-20

Subject Terms: Violence—forecasting; violence—risk assessment; forensic psychiatry; dangerously mentally ill—risk assessment; criminals—risk assessment

ACKNOWLEDGEMENTS

I am grateful to many individuals who played a substantial role in helping me arrive at this point. Most obviously, I am tremendously thankful for having had the opportunity to learn from and work with teachers who I admire and respect so greatly. I am indebted to Kevin Douglas for his mentorship and friendship over the past several years. His thoughtfulness, patience, and kindness were important sources of support; his high standards coupled with his generosity in providing many professional opportunities made for truly gratifying doctoral years. I am privileged to be his first Ph.D. graduand.

John Edens has been a role model to me since I started graduate school, and is responsible for fostering my interest in the field of violence risk assessment. He too has provided innumerable opportunities for me, and I am profoundly grateful both for his camaraderie and the tremendous hard work, time, and energy he has invested in my professional development.

I am very thankful to Stephen Hart for his lessons in pragmatics, willingness to give his time and knowledge to help me work through statistical quandaries, and help in seeing problems from fresh perspectives. He has been a consistent, gracious, and friendly source of support.

I also am appreciative of several other individuals who provided guidance to me during my graduate training. Mary Alice Conroy taught me the fundamentals of forensic assessment when I started graduate school at Sam Houston State University, and she instilled in me a genuine love of forensic clinical practice. The sensitivity and "humanness" that my internship supervisors, Jack Terry and Bill Warnken, showed me this past year were extraordinary.

In the process of completing this project, I have accumulated a long list of persons to whom I am indebted. Several organizations provided funding for this research: a Social Sciences and Humanities Research Council of Canada Doctoral Fellowship, a Dissertation Award from the American Academy of Forensic Psychology, and an American Psychology-Law Society (Division 41 of the American Psychological Association) Student Grant-In-Aid. I also received aid through SFU's Work-Study Program to hire several research assistants who assisted with article retrieval. I am grateful to the extremely capable staff at SFU's Interlibrary Loan Program, especially Vera Yuen, who put in an astounding amount of time during a period when the scope of the present project was considerably larger to track down documents that were decades old or

otherwise difficult to locate. I also thank Michelle Collins, Diane Strub, Claudia Pace, and Robin Simcoe, who assisted in screening a portion of the literature searches. More generally, I have Michelle to thank for being my stand in for various administrative and research tasks while I was away on internship and not able to be present in person.

Finally, I appreciate tremendously the efforts made by numerous researchers around the world who very graciously responded to my requests for copies of unpublished disseminations or additional information needed to code methodological details or compute effect sizes: C. Allen, L. Augimeri, P. Bartel, H. Belfrage, R. Borum, C. de Ruiter, S. Desmarais, M. Dolan, G. Garman, A. Gibas, M. Grann, B. Howard, P. Kennealy, R. Kropp, K. Levene, W. Lindsay, H. Lodewijks, L. McKinnon, E. Nicholson, S. Penney, K. Reeves, H. J. M. Schönberger, J. Skeem, C. Stadtland, J. Viljoen, G. Vincent, K. Watt, and J. Worling. Many other individuals provided additional disseminations, which are included in a paper in preparation, after the cutoff date had passed for coding studies in the dissertation.

Without doubt, life during graduate school would have felt near hardship in the absence of the friendships I had before and those I made along the way, especially the FCPSO gang at SHSU, the ladies in the CFL and MHLPI at SFU, and internship "journal club" members at UMass Med School. Finally, I couldn't possibly imagine being gifted with a more encouraging, loving and supportive family - my mom and dad, sisters, brother, and brother-in-law were there for me all the way, no holds barred.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Professionals who evaluate individuals' risk for engaging in future violent acts have numerous instruments or tools at their disposal to assist them in this enterprise. Generally, such instruments align either with the actuarial, or statistical, approach or with the empirically-guided clinical, or structured professional judgment (SPJ), approach. Whereas instruments based on the former method have been available for several decades (e.g., Burgess, 1928; Glaser, 1954, 1955a, 1955b), and continue to be developed today, the SPJ approach became nascent only during the past decade and a half. During that time, several "SPJ tools" have been developed, and the SPJ approach has been introduced into standard clinical practice in numerous correctional and psychiatric settings in North America and abroad.

Although many individual studies of various SPJ tools have been completed, and circumscribed meta-analytic inquiries comparing the two general approaches have been carried out (e.g., Campbell, French, & Gendreau, 2007; Hanson, Helmus, & Bourgon, 2007; Hanson & Morton-Bourgon, 2004, 2007), the empirical foundation of the SPJ approach to date has not been examined comprehensively. Moreover, claims made by proponents of the actuarial approach that the SPJ (or any non-actuarial) model is inferior to actuarial methods under any and all circumstances (e.g., Rice, Harris, & Hilton, in press) cannot be evaluated properly in the absence of a clear understanding of how the SPJ approach, as well as specific SPJ tools, performs. As such, the primary purpose of this dissertation is to evaluate the state of the empirical research on the predictive validity of the SPJ approach and each of the specific measures developed within its framework. An ancillary aim of the project is to examine the performance of the SPJ approach when compared directly to other types of prediction approaches. Given the clear benefits that quantitative synthesis offers over narrative literature reviews in terms of consistency and

transparency (see, e.g., Lipsey & Wilson, 2001; Rosenthal & DiMatteo, 2001), meta-analytic techniques are applied in this dissertation to evaluate all available research regarding the predictive validity of the SPJ model.

In the sections that follow, a brief overview of the legal relevance of violence risk assessment will be presented and the most common approaches to evaluating risk will be delineated, including a more detailed examination of the contours that define that the SPJ approach. Descriptions of the 16 SPJ tools developed to date (12 for use with adults, two for adolescents, and two for children) will be provided, although the specific nature of the empirical support for each measure will be examined quantitatively in the Results section. Finally, based on the rationale provided throughout this Introduction, the section will conclude with the identification of specific research questions that will be investigated.

Violence Risk Assessment

The assessment and management of risk for violence to others is a clinical task that mental health professionals perform regularly in virtually every psychiatric setting. Moreover, more formal violence risk assessments are required in many legal and quasi-legal situations (Lyon, Hart, & Webster, 2001; Shah, 1978), such as civil commitment, bail determination, juvenile transfer and decertification, and release decision-making (hospitals, correctional facilities). Lyon et al. (2001) identified 17 disparate areas raised by Canadian law in which assessments of risk for violence are required. In addition to the more obvious areas such as criminal, civil, and family law, violence risk assessment also is required in workers compensation law, immigration law, freedom of information/privacy law, and national security efforts.

The expanding role of violence risk assessment in the practice of clinical psychology and psychiatry over the past several decades (see Douglas, Macfarlane, & Webster, 1996; McNiel, 1998) has resulted in part from a series of developments in the legal arena. Whereas the basis for

involuntary hospitalization was once "need for treatment," this criterion was replaced in the 1960s by "dangerousness to others." In the 1972 case *Lessard v. Schmidt*, "dangerousness" was introduced into civil commitment statutes as a basis for commitment in the United States (U.S.) Also during the 1970s, the practice of imposing tort liability on clinicians who negligently failed to protect potential victims was established (*Tarasoff v. Regents of the University of California*, 1976). During the following decade, in the 1980s, many states enacted statutes authorizing involuntary treatment in the community for otherwise "dangerous" patients. In the 1990s, risk assessments of violence were mandated explicitly in the Americans with Disabilities Act (1990), which protects the employment rights of people with disabilities unless those disabilities result in an employee posing a "direct threat" of violence to co-workers or customers.

In Canada, the expansion of risk assessment primarily has resulted from legislative developments (see Lyon et al., 2001). The federal government introduced two legislative measures during the 1940s to deal with chronic and dangerous offenders. The habitual criminal legislation was enacted in 1947, and the criminal psychopath legislation was proclaimed the following year. These two initiatives laid the groundwork for the dangerous offender legislation, which became law in 1977 (Part XXIV, Criminal Code; see Lyon et al., 2001). Twenty years later, in 1997, Bill C-55 was enacted, which included significant amendments to Part XXIV of the Criminal Code and was designed to deliver more severe sanctions to repeat and violent offenders. One of the Act's provisions was the creation of the Long-term Offender (LTO) designation. This designation is given to offenders who have been convicted of a serious personal injury offence (as defined by the Criminal Code) and who are judged as being likely to re-offend. The LTO legislation was developed primarily to address concerns that additional attention needed to be given to serious sexual and violent offenders who fell short of meeting criteria under the Dangerous Offender legislation. Cases in Canada (*Re Moore and the Queen*, 1984) and the United States (*Barefoot v. Estelle*, 1983; *Kansas v. Hendricks,* 1997; *Schall v. Martin*, 1984) have

legitimized the role of mental health professionals (and in some cases their reliance upon clinical predictions of violence) as risk assessors.

The costs associated with errors in risk assessments are serious because such evaluations implicate important social values such as protection of the public and individual civil liberties (see Schopp, 1996). Individuals who are assessed incorrectly as being at high risk for violence may face involuntary civil commitment, mandated outpatient treatment, forced administration of medication, and loss of a host of civil liberties (Canadian Charter of Rights and Freedoms, 1982; *Lyons v. the Queen*, 1987; see also Monahan et al., 2001b). Individuals who are assessed incorrectly as not being at risk for violence, and hence permitted to retain their freedom, may violate others' civil rights and legally protected interests to be free from injury by others. In light of the serious consequences that violence risk assessment may influence, the importance of performing such evaluations according to the highest standard of clinical practice is self-evident.

Not surprisingly, over the years, the task of assessing risk for violence to others has been conceptualized and defined in various ways. One group of prominent researchers regards risk assessment as the establishment of a probability of violence along an ordered categorical distribution (Quinsey, Harris, Rice, & Cormier, 2006). As such, within this perspective, the assessment task focuses on forecasting or predicting violence. For clinicians in applied contexts, however, this unduly narrow conceptualization of violence risk assessment is unhelpful. A relatively different view of the clinical task identifies violence risk assessment as a two-part process that entails: (a) evaluating an individual to characterize the risk that she or he will be violent in the future and (b) developing a contingency-based action plan to manage or reduce that risk (Hart, 1998, 2001). The intentional benefit of a violence risk assessment within this perspective is that, if conducted carefully, the assessment has potential to evaluate and manage risk of future violence and thus can play a critical role in preventing violence (Douglas, Cox, & Webster, 1999; Douglas & Kropp, 2002) by identifying persons at risk of violence and offering or

compelling them to receive mental health treatment or other services (Buchanan & David, 1994; Cole & Glass, 2005).

Approaches to Assessing Risk for Violence

Various ways of assessing risk have been discussed in the literature, and each can be understood as falling on a continuum in terms of the degree of structure imposed on the three central decisions to be made in the assessment process: 1) which risk factors to consider and how to measure them; 2) how to combine the risk factors; and 3) how to generate the final risk estimate (Monahan, 2008). Stemming from Meehl's (1954) demarcation, the two major processes used in assessing risk for future violence are *actuarial* and *clinical* (also see Douglas & Kropp, 2002; Grove & Meehl, 1996; Melton, Petrila, Poythress, & Slobogin, 2007). The *actuarial* approach typically involves reaching judgments based on statistical information according to fixed and explicit rules. Grove and Meehl (1996, p. 293) defined actuarial prediction as one that "involves a formal, algorithmic, objective procedure (e.g., equation) to reach the decision." Usually, this entails assigning a score to a limited number of weighted and predetermined predictors, summing them, and interpreting the sum of the risk factors "as an 'actuarial' graduated probability measure, representing the amount of risk attributed to the individual" (Doyle & Dolan, 2002, p. 651). Stated another way, weighted items are combined using a predetermined, numerical weighting system (i.e., an algorithm or equation) to yield a decision (Hanson, 1998). Structure is imposed on each of the three major decisions in the actuarial assessment approach: there is no discretion in terms of selecting, measuring, or combining risk factors, and the final risk estimate is determined by a priori, fixed rules.

There are various types of clinical approaches (referred to as professional judgment approaches by Hart et al., 2003), each of which also can be conceptualized in terms of the degree of structure imposed on the decision-making process: unstructured, anamnestic, and structured.

Decisions that are justified solely on the basis of an evaluator's clinical experience and/or qualifications have been referred to as *unstructured professional judgment* (Hart, 2001; Hart et al., 2003) or *unguided clinical assessment* (Hanson, 1998). In the unstructured clinical approach, no structure is imposed on any of the three major decisions in the assessment process. The *anamnestic* approach to risk assessment (Hart et al., 2003; Melton et al., 2007; Miller & Morris, 1988; Otto, 2000) is based on clinical decision-making that has some degree of structure (in terms of providing guidance as to the nature of the risk factors that should be considered), and entails considering an individual's life history to determine which personal and situational risk factors were associated with violence in the past and then determining if these factors still exist.

The third type of clinically based approach to violence risk assessment is *structured professional judgment* (SPJ) (Douglas et al., 1999; Hart, 1998; Webster, Douglas, Eaves, & Hart, 1997), also referred to as structured clinical judgment, structured discretion, or empirically-guided clinical judgment (Hart, 2008). The SPJ approach was developed to provide a method for assessing violence risk that minimized the perceived limitations of both unstructured clinical judgment and actuarial approaches. In line with recommendations to use more structured procedures to optimize the accuracy of violence risk assessments (e.g., Borum, 1996), this clinical-based decision-making model is defined by the use of guidelines "that reflect current theoretical, clinical, and empirical knowledge about violence" (Douglas & Kropp, 2002, p. 626). Such guidelines are grounded in discourse from the scientific, professional, and legal literatures. In this approach to assessing risk for violence, structure is imposed on which risk factors should be considered and how they should be measured, but the way in which factors are combined is left to the discretion of the evaluator. The evaluator's discretion similarly is valued in terms of generating the final estimate of risk. The SPJ approach is described in further detail below.

In general, research in the violence risk assessment field has shifted from studying clinicians' accuracy (although see Skeem, Mulvey, et al., 2005; Skeem, Schubert, et al., 2005) to

developing risk assessment technologies to improve it. This shift seems to have occurred in concert with criticisms that unstructured clinical judgment was associated with unacceptably poor levels of reliability and validity (Dawes & Corrigan, 1974; Dawes, Faust, & Meehl, 1989; Lidz, Mulvey, & Gardner, 1993; McNiel & Binder, 1995; Meehl, 1954; Monahan, 1981). A notable exception to this shift away from studying clinicians' accuracy was Mossman's (1994) meta-analysis on the topic, in which he demonstrated how evaluations of diagnostic accuracy can be compromised when accuracy is measured with techniques that are affected by base rates or evaluators' preferences for certain outcomes (i.e., a tendency for one's inevitable mistakes to be more likely to comprise false positives or false negatives).

Following a meta-analytic review that used effect sizes generated from receiver operating characteristics (ROC) analyses, Mossman advanced more positive conclusions regarding the accuracy of clinically based predictions compared to the pessimism regarding clinicians' ability to predict violence that prevailed at the time (e.g., Cocozza & Steadman, 1976; Ennis & Litwch, 1974). More specifically, examining 58 data sets from 44 published studies of violence prediction, Mossman concluded that clinicians could distinguish violent from nonviolent patients with a modest, better-than-chance level of accuracy. The mean weighted area under the curve[1] (AUC) of cross-validated discriminant functions (i.e., actuarially based predictions) was .71 (SE ± 0.01), which was greater than the mean weighted AUC for clinical predictions (.67, SE ± 0.01; $z = 2.89$, $p = .004$, two-sided). Of note, however, Mossman reported that this difference could be explained by the relative accuracies of short vs. long-term (> 1 year) predictions, such that the average accuracy of cross-validated discriminant functions covering less than 1 year was comparable to the average for clinical predictions. Despite Mossman's report of relatively larger indices of accuracy than had been observed in the literature to that point, there remained clear

[1] See page 60 for a detailed explanation of this statistical index.

boundaries to the accuracy that could be achieved by unstructured clinical approaches, which was an impetus to developing tools according to the SPJ framework.

The Great Debate: A Brief Synopsis

For decades, there has been vigorous commentary in the research and professional literature regarding the alleged superiority of the actuarial risk assessment approach in general, as well as with respect to violence risk assessment specifically (e.g., Arnhoff, 1960; Buchanan, 1999, 2008; Dawes, 2005; de Groot,1961; Gendreau & Goggin, 1996; Gottfredson & Gottfredson, 1988; Gottfredson & Moriarty, 2006a, 2006b; Grebstein, 1963; Grubin, 1997; Harris & Rice, 2007; Harris, Rice, & Quinsey, 2008; Heilbrun, Douglas, & Yasuhara, in press; Hilton, Harris, & Rice, 2006; Hirsch, 1972; Holt, 1970, 1986; Kahn, 1960; Lindsay & Beail, 2004; Litwack, 2001; Mills, 2005; Mulvey, Lidz, Gardner, & Shaw, 1996; Richards, 1963; Sarbin, 1943, 1962; Shah, 1978; Sines, 1970; Stricker, 1967; Thorne, 1960). Although some researchers recommend that clinical judgment be supplanted completely by statistical devices (e.g., Grove & Meehl, 1996; Harris, Rice, & Cormier, 2002; Quinsey et al., 2006), others have regarded such a strategy as an unfounded recommendation lacking empirical support (see Hart et al., 2003; Litwack, 2001; Skeem, Schubert, et al., 2005).

At least two meta-analyses focusing on the relative performance of decisions made using clinical and actuarial methods have been completed (additional meta-analyses, reviewed below, focused specifically on predictions about violence). Most recently, Ægosdóttir et al. (2006) examined 92 effect sizes coded from 67 studies in which predictions for a range of outcomes made using statistical approaches were compared directly with mental health professionals'

clinical predictions. The authors reported a small effect in support of the superiority of statistically based predictions, with the effect size of difference $d = -.15$.[2]

In an earlier study, Grove, Zald, Lebow, Snitz, and Nelson (2000) presented a meta-analytic review of the general decision-making literature. Among 136 direct comparisons between actuarial and clinical approaches, 46% favoured the actuarial approach; 46% yielded no advantage for either approach; and 8% favoured the clinical approach. The magnitude of these differences, however, was not great. As noted by Hart et al. (2003), among the direct comparisons in which the actuarial approach was superior, the increment in accuracy was only about 10%. Hart et al. (2003) highlighted two additional important points regarding Grove et al.'s quantitative synthesis. First, the majority of the studies in which direct comparisons were made between the two decision-making approaches did not involve predictions or prognoses about the future, but rather involved decisions or diagnoses about present state, which is noteworthy in that predicting the future ostensibly is more difficult than "predicting" the present. Second, very few studies compared actuarial and clinical forecasts regarding violence. One year following Grove et al.'s meta-analysis, Litwack (2001) identified only eight studies[3] in which such direct comparisons for judgments about violence risk were performed, none of which offered clear evidence of the superiority of either decision-making strategy.

Set against the backdrop of this well-entrenched debate has been the development of an astoundingly large number of tools or measures that purportedly assess risk for violence to others. In fact, a comprehensive review of the literature (both published and unpublished) that has accrued over the past five decades indicated the existence of at least 457 tools (Guy, 2008). Examining a subset of these tools, Campbell et al., (2007) completed a meta-analysis in which

[2] Negative value indicates superiority of statistical prediction.

[3] Gardner, Lidz, Mulvey, and Shaw (1996); Glaser (1955b, 1962); (Hall, 1988); Hassin (1986); Holland, Holt, Levi, and Beckett (1983); Quinsey and Maguire (1986); Smith and Lanyon (1968); and Wormith and Goldstone (1984).

they compared the predictive validity (for violence) of various self-report, actuarial, and structured clinical risk protocols. The studies included in analyses were restricted to those that were prospective in nature and sampled only adults who were offenders or forensic psychiatric patients; the outcome of analysis was restricted to nonsexual violence. The authors failed to find evidence of the overall superiority of the actuarial approach, and concluded: "no one measure stood out as the most effective for predicting violent recidivism" (Campbell et al., 2007, p. 20). Although SPJ tools were included in analyses, the authors did not evaluate the performance of the tools as they were intended to be used in clinical practice (i.e., summary risk ratings were not included). As such, although informative, the study's findings are limited in terms of their ability to evaluate fully the performance of the SPJ approach to violence risk assessment.

In a meta-analysis that focused on the validity of risk assessments for intimate partner violence, Hanson et al. (2007) compared predictions coded from 18 non-redundant samples that were based on four approaches: spousal assault scales, other risk scales, structured professional judgment, and victim judgment. The four approaches performed similarly (and were not significantly different from one another) with respect to predictions for spousal assault recidivism.

In contrast to the meta-analytic results presented by Campbell et al. (2007) and Hanson et al. (2007), Hanson and Morton-Bourgon (2004), in their meta-analysis of 95 samples of sexual offenders, found actuarial risk instruments to evidence predictive superiority relative to other types of approaches (as was found in an earlier quantitative review on the topic; Hanson & Bussiére, 1998). Hanson and Morton-Bourgon (2004, p. 17) concluded, "Empirically-guided professional judgments showed predictive accuracies that were intermediate between the values observed for clinical assessments and pure actuarial approaches. The same pattern of results applied to the prediction of sexual recidivism, violent non-sexual recidivism, and general (any) recidivism." Depending on the specific comparison, however, the difference in magnitudes

fluctuated (or disappeared). For example, pertaining to sexual recidivism, the mean effect size (when the outlier was retained) for empirically guided and general (i.e., not specifically for sexual offending) actuarial tools was the same: $d = .51$.

In a more recent study, Hanson and Morton-Bourgon (2007) cumulated findings from 100 studies of offenders released following an index sexual offence. When the predictive accuracy of risk assessments based on structured professional judgements were compared to assessments based on actuarial measures and unstructured professional judgement, no statistically significant differences were observed between the three approaches for the prediction of sexual recidivism. The largest effect size for sexual recidivism was for structured professional judgments made using the Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997); the mean Cohen's *d* (Cohen, 1988) was 1.11 (Mdn $d = 1.23$). Although the authors noted, with merit, that this finding was based on only three studies, and that significant variability was present ($Q = 7.96$, $df = 2$, $p < .05$), many actuarial tools based on only a few additional studies with comparable degrees of variability yielded much smaller effect sizes for the prediction of sexual recidivism (e.g., Statistical Index of Recidivism (Nuffield, 1982): $k = 4$, mean $d = .52$, Mdn $d = .79$; $Q = 10.61$, $df = 3$, $p < .05$; Violence Risk Appraisal Guide (Harris, Rice, & Quinsey, 1993); $k = 7$, mean $d = .51$, Mdn $d = .46$; $Q = 8.23$, $df = 6$, $p =$ n.s). The only SPJ tool included in this meta-analysis was the SVR-20, which was developed to assess risk for sexual recidivism. The SVR-20 performed comparatively worse when 'any violence' and 'any reoffending' were the outcomes under consideration (i.e., compared to actuarial measures and compared to its performance for predicting sexual violence). Across all types of tools, in fact, risk judgments were most accurate when used to predict the type of outcome for which they were designed to be applied.

Taken together, the meta-analytic evidence presented thus far indicates that there often is no definitive advantage - in terms of predictive accuracy—for either actuarial or structured

clinical approaches. When differences are observed, they most often are in the direction toward the actuarial approach; the magnitudes of the differences, however, typically are small. Therefore, given no clear advantage for either approach as far as predictive validity is concerned, factors other than predictive accuracy indices should shape practitioners' decisions regarding which risk assessment strategy or approach to follow. Lavoie, Guy, and Douglas (in press) presented five characteristics or principles that a defensible risk assessment procedure would embody, and which therefore would be important to consider when selecting an instrument for clinical practice.

Principles Guiding Selection of an Assessment Paradigm

First, a risk scheme should include relevant scientifically supported risk factors. Decades of research have yielded important evidence regarding which specific factors demonstrate robust empirical associations with antisocial behaviour. In fact, a sufficiently large number of studies has been completed to warrant quantitative syntheses within divergent areas of the literature, including mentally disordered offenders (Bonta, Law, & Hanson, 1998); sexual offenders (Hanson & Morton-Bourgon, 2005); stalkers (Rosenfeld, 2004); prisoners (Gendreau, Goggin, & Law, 1997); adult offenders (Gendreau, Little, & Goggin, 1996); juveniles (Cottle, Lee, & Heilbrun, 2001; Heilbrun, Lee, & Cottle, 2005); and delinquent girls (Hubbard & Pratt, 2002). Moreover, in addition to including appropriate risk factors, the risk scheme must operationalize the factors clearly. Such semantic precision is critical both to promote inter-rater reliability as well as to remind users of the empirical association between violence and the risk factors included in the scheme.

Second, a good risk scheme has comprehensive coverage of violence risk factors. An evaluator who uses a scheme that does not contain most identified risk factors may render inappropriately low or high estimates of risk status should potentially important and pertinent risk factors exist for a case that were not considered.

Third, a risk scheme should facilitate the construction of a plan that shows promise in reducing risk, and provide guidance regarding strategies to manage such risk reduction efforts. To accomplish this task, the tool must contain dynamic violence risk factors that represent targets for treatment (Douglas & Skeem, 2005). Such risk factors also must be criminogenic (i.e., changing the degree to which a factor is present for a given individual will result in an attendant change in the individual's level of risk for violence; see, generally, Andrews & Bonta, 2006).

Fourth, a risk scheme should provide a clear and logical method of communicating risk decisions. Specifically, the scheme should articulate the degree of risk posed and should detail the specific actions to reduce the risk. Regarding the presentation of degree of risk, the scheme should allow the assessor to offer categorical, descriptive risk statements tied to the estimated risk level and to the associated intensity of management anticipated to reduce risk. An alternative method involves using numerical or probabilistic estimates of risk based on the presence of risk factors that are valid for a particular length of time. Sole reliance on this mode of presentation is inappropriate because specific prescriptive actions are not aligned with numerical or probabilistic estimates; therefore, opportunity to reduce risk is not present.

Finally, because consequences of violence risk assessments can be serious, as described above, the decision-making process should be entirely transparent and available for review by relevant stakeholders. Transparency in the risk assessment process entails a clear statement of risk factors included, scoring rules, and principles followed to integrate risk factors into a final risk decision.

If one endorses the view that risk prediction is relatively meaningless in the absence of efforts to prevent, reduce, and manage risk, and given the nature of the principles that Lavoie et al. (in press) recommended, the approach to violence risk assessment espoused by the SPJ model is a defensible decision-making framework to examine. Indeed, international practice guidelines support the use of the SPJ approach. For example, after conducting a thorough review regarding

the extent to which tools developed to assess violent, sexual, and general recidivism have been empirically validated, the Standards and Guidelines published by Scotland's Risk Management Authority (RMA) arrived at the following conclusion:

> When conducting a detailed risk assessment for any offender, the use of tools based upon static assessment items—actuarial tools—is permissible only when it forms part of a structured professional assessment; identifying risk and protective factors specific to the individual; and formulating risk in an analytical manner. This is due to the limitations of actuarial tools in the crucial tasks of both identifying risk and protective factors specific to the individual and also guiding practitioners in the formulation of risk leading to tailored risk management plans. (RMA, 2007, p. 7)

Similarly, in a document that presents practice guidelines for the assessment and management of risk to self and others in mental health services prepared for England's National Mental Health Risk Management Programme (Department of Health, 2007), Best Practice point 10 reads, "Where suitable tools are available, risk management should be based on assessment using the structured clinical judgement approach" (p. 18; see p. 20 for a detailed explanation of this point).

Moreover, in addition to the many reasons for adhering to the SPJ approach that are directly relevant to clinical practice, there also is a legal impetus. As noted by Kropp and Hart (2004)[4]: "The Supreme Court of Canada, in considering a wide range of cases related to violence and violence risk over many decades, has consistently held that the application of discretion by criminal justice and mental health professionals (e.g., police and corrections officers, prosecutors and judges, parole and review boards, psychiatrists and psychologists) is both necessary and appropriate."

[4] Kropp and Hart provided several relevant legal citations: Canadian Foundation for Children, Youth and the Law v. Canada (Attorney General), [2004] 1 S.C.R. 76; Penetanguishene Mental Health Centre v. Ontario (Attorney General), [2004] 1 S.C.R. 498; R. v. Johnson, [2003] 2 S.C.R. 357; Smith v. Jones, [1999] 1 S.C.R. 455.

The SPJ Approach

Much has been written describing the SPJ approach to violence risk assessment (e.g., Bloom, Webster, Hucker, & De Freitas, 2005; Douglas, Cox, et al., 1999; Douglas & Kropp, 2002; Douglas & Ogloff, 2003a, 2003b; Douglas, Ogloff, & Hart, 2003; Webster, Müller-Isberner, & Fransson, 2002). A defining characteristic of a SPJ measure (see below for reviews of individual measures) is that it comprises guidelines, or an *aide-mémoire*, developed to assist evaluators in exercising discretion during the process of conducting a violence risk assessment. Another defining characteristic is that SPJ guidelines are constructed with input from various groups of stakeholders, such as researchers, clinicians, and administrators.

Concretely, SPJ guidelines are checklists that comprise factors that have demonstrated an empirical association with increased risk for violence. Items have a unit-weighting scheme, and the decision-maker weights the items. Risk is presumed not to be a simple linear function of the number of risk factors present in a case, although it is generally true that risk is greater when relatively more risk factors are present. Of course, an additive model is overly simplistic, and, in fact, critical to the foundation of the SPJ approach to assessing risk is the presupposition that a single risk factor may play a disproportionate role in increasing an individual's level of risk (see also Monahan et al., 2001b).

Recently, Hart (2008) noted that guidelines developed within the SPJ framework are distinguished in terms of being preventive, structured, and flexible. Schemes are preventive (i.e., not only predictive) in that they guide evaluators first to consider the individual and contextual factors that are believed to increase or decrease risk for violence, and, second, to identify specific interventions that may be useful for managing and/or reducing that risk. Importantly, they also assist in communicating decisions about risk in part by offering general suggestions pertaining to the practices of making and communicating decisions about risk.

SPJ schemes are structured in that they provide explicit recommendations pertaining to the processes of identifying risk factors, combining information about the relevance of the factors for the individual under assessment, and communicating the information. By definition, therefore, the SPJ approach is incompatible with risk judgments that are based solely on intuition. SPJ schemes also are structured in that they organize activity, reduce cognitive burden, and facilitate pattern recognition. As a result of the structure imposed on the assessment, decisions are more consistent and the decision-making process is more transparent than they otherwise would be, thereby enhancing accountability.

SPJ schemes are flexible in the sense that assessments are individualized and contextualized. In contrast to imposing fixed, explicit algorithms, the evaluator's discretion is valued. Moreover, an SPJ scheme contains factors that are expected to change over time in saliency and relevance in different ways for different individuals. More generally, because they are based on current discourse in the empirical, professional, and legal spheres, SPJ schemes also are flexible in the sense that they may, for the most part, be considered as "works in progress." Indeed, as observed by Webster et al. (2002, p. 191): "No clinical assessment device, certainly not one in the complex area of risk assessment, should ever be considered fixed or formalized for use into the indefinite future."

As described by Borum, Bartel, and Forth (2003, p. 4), "The structured professional judgment approach helps to focus the evaluator on relevant data to gather during interviews and record reviews, so that the final judgment, although not statistical, is well informed by the best available research." Summary judgments are "global or conclusory opinions about severity of risks and degree of effort or intervention required to manage risk" (Hart, 2008). Although some SPJ manuals encourage users to communicate such decisions about risk in a categorical manner (low, moderate, or high), this practice is neither a necessary nor defining feature of the SPJ

approach.[5] Indeed, research regarding the use of other formats is underway. For example, initial results from an investigation using a 5-level format for communicating final decisions about risk level using the Structured Assessment of Protective Factors (SAPROF; de Vogel, de Ruiter, Bouman & de Vries Robbé, 2007, as cited in de Vries Robbé, de Vogel, & de Spa 2008) and HCR-20 are encouraging (de Vries Robbe et al., 2008).

Irrespective of format, summary risk ratings are intended to be helpful for prioritizing cases. For example, using the familiar tri-categorization scheme, a decision that an individual's level of risk is high indicates that several risk factors (or at least one highly significant factor) are present and relevant for the individual and that, in the absence of appropriate, individualized intervention, the potential for the individual to act violently is increased.

SPJ Risk Assessment Tools

During the past decade, several measures or decision support tools consistent with the SPJ approach have been created and examined empirically. Recognizing the importance of human development across the lifespan, various SPJ schemes are available for use with adults, adolescents, and children. Additionally, schemes that assess risk for both general violence and specific forms of violence have been developed.

The volume of research completed for each measure varies (including some for which investigations of predictive validity have not yet been completed), as do findings regarding predictive validity, with results from some studies offering support for the validity of these tools, and others whose results are less favorable. Given the variability in research findings, the growing use of the SPJ approach to violence risk assessment as noted above, as well as the potentially

[5] Although this form of categorization, as noted, is not necessary to practice within an SPJ framework, research suggests that clinicians prefer categorical risk communication that is management-oriented and prescriptive in nature, compared to probability or numeric estimates of risk (Heilbrun, O'Neill, Strohman, Bowman, & Philipson, 2000).

serious consequences involved in any such evaluation, it is important to take stock of the empirical work that has been completed to date in order to inform both practice (is it appropriate to use certain SPJ tools under certain circumstances, or even at all?) and science (what gaps exist in the empirical knowledge base regarding SPJ tools and their application?). In this section, all SPJ schemes known to the author that have been developed to date will be described. Empirical findings for each tool whose predictive validity has been studied will be presented quantitatively in the Results section.

Tools for Use with Adults: General Violence

Historical-Clinical-Risk Management-20 (HCR-20; Webster et al., 1997)

Initially published in 1995 (Webster, Eaves, Douglas, & Wintrup, 1995), the HCR-20 was the first SPJ measure developed to assess general violence among adults. The HCR-20 has been well-researched since its initial publication, with approximately 85 studies now available that have investigated its reliability or predictive validity (for an annotated bibliography, see Douglas, Guy, & Weir, 2007; for reviews of the tool, see Buchanan, 2001; Mossman, 2000; and Witt, 2000).

The HCR-20 is so-named for its inclusion of 20 risk factors in Historical, Clinical, and Risk management domains. The instrument contains 10 historical, largely static risk factors such as psychopathy, previous violence and past substance use problems and 10 potentially changeable, dynamic risk factors (see Appendix 1.1)[6]. Five of these concern current mental and clinical status such as negative attitudes and active symptoms of major mental illness (the Clinical scale), and five concern future situational risk factors such as lack of plan feasibility and treatment noncompliance (the Risk Management scale). Because they are sensitive to change, the

[6] Items of the SPJ tools included in the analyses will be presented in Appendixes.

dynamic risk factors comprising the Clinical and Risk Management scales are intended to facilitate development of risk management and intervention plans (Belfrage & Douglas, 2002).

Consistent with the SPJ model, items were selected on the basis of their association with violence in the scientific and professional literatures available at that time, as well as by consultation with forensic mental health professionals. Each item is rated on a 3-point scale, with 0 indicating that *available information contraindicates the presence of the item*, 1 indicating that *available information suggests the possible presence of the item*, and 2 indicating that *available information indicates the presence of the item*. By design, numerical cut-off scores or algorithms that mandate categorization of individuals into risk levels are not provided, although the user can sum items on each of the HCR-20's three scales to generate numerical scores. Rather, clinicians are encouraged to communicate level of risk using categorical levels of low, moderate, and high. Such estimates are based on: (a) the assessment of the risk factors, (b) the relative importance for a given individual, and (c) the degree of intervention estimated to be necessary to prevent violence.

Because items on an SPJ tool are intended to reflect broadly the current state of scientific and professional literature, such schemes will require periodic updating. The HCR-20 was published just over a decade ago; since then, thousands of studies on violence have been published. For that and other reasons, a revised version of the HCR-20 is being developed. As one part of the development of the *HCR-20, Version 3* (Douglas, Hart, Webster, Belfrage, & Eaves, in preparation), a comprehensive evaluation of the research and professional literature published since 1997 was conducted (Guy & Wilson, 2007) to inform decisions regarding the continued importance of the various risk factors and to identify ways in which the operationalization of the factors potentially may be refined to reflect current research. A large-scale empirical project currently is underway to examine inter-rater reliability, structural reliability, internal consistency, predictive validity, incremental validity, and item functioning of

the HCR-20 to provide part of the empirical foundation for Version 3 (e.g., Guy & Douglas, 2006; Douglas, et al., in preparation; Douglas & Lavoie, 2006).

In addition to the current item level coding system (i.e., 0, 1, and 2), the forthcoming version of the HCR-20 will evaluate an expanded coding range that will include an option for assigning a score of 3 to indicate that a variable is 'present and extreme.' Additionally, an indicator system is being tested wherein, for each risk factor, a set of potential manifestations of the item (*not present, possibly present,* or *definitely present*) will be offered to facilitate ratings regarding the severity and individual manifestation of the risk factors. Ratings regarding the idiographic relevance of each item (*not relevant* or *relevant*) also will be included.

The HCR-20 has been integrated formally into several forensic and civil systems in North America, such as the Correctional Service of Canada; the Ohio Department of Mental Health; New York Office of Mental Health, Forensic Bureau; and the Forensic Services Division of the Department of Mental Health in California. It also has been implemented in individual forensic, civil, or correctional facilities in many countries, including in Sweden, Germany, Australia, Japan, Canada, the United States, the United Kingdom, Norway, the Netherlands, and Demark (personal communication, K. Douglas, August 23, 2008).

The Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Middleton, 2004)

The START:

> is a structured clinical scheme to organize assessments, guide clinical interventions, and index possible improvement due to therapeutic interventions and other events, as well as map any evidence of mental and behavioral deterioration, across seven often-overlapping risk domains: risk to others, self-harm, suicide, self-neglect, substance abuse, unauthorized leave, and victimization by others (Nicholls, Brink, Desmarais, Webster, & Martin, 2006, p. 314)

The scale comprises 20 dynamic risks and strength-related factors, and uses two separate scales for strengths and risks (see Appendix 1.2).[7] Items on both scales are scored in a manner consistent with the HCR-20 (i.e., 0 = item is not evident, 1 = item is evident to some extent, 2 = item is present). Raters additionally are encouraged to designate particular items as being "critical" risks or "key" strengths, as appropriate; these items are intended to be used to develop a risk management plan. Finally, in addition to scoring individual items, raters make Summary Risk Estimates (coded 0 = low, 1 = moderate, 2 = high) across the seven domains. In a user satisfaction study (see Webster et al., 2004), psychiatrists, nurses, social workers required on average 8 minutes to complete the START. To date, only one empirical evaluation of the START's predictive validity has been published in a peer-review outlet (Nicholls et al., 2006).

Structuring Clinical Judgment: Risk (SCJ: Risk; Hogue & Allen, 2006)

An adaptation of the HCR-20, the SCJ: Risk (see Appendix 1.3) was developed to meet the security needs of a high-secure hospital in the United Kingdom (U.K.). Following a number of high profile incidents, a review of security (the 'Tilt' review; Tilt, Perry, Martin, Maguire & Preston, 2000) recommended the implementation of a system to document risk decisions with particular reference to: (a) the immediate risk of harming others; (b) the risk of suicide or self-harm; (c) vulnerability to risk from others; (d) the risk of escape; and (e) the risk of organized action in collaboration with others to subvert security and safety. The review highlighted the importance of integrating the risk assessment and management processes into routine clinical practice, citing the importance of evidence-based practice and multidisciplinary collaboration.

The SCJ: Risk includes all 20 of the HCR-20's items and retains the same format of item grouping (historical, clinical, and future-oriented factors). An additional 10 items are included on

[7] See Webster, Nicholls, Martin, Desmarais, and Brink (2006) for an overview of the precursor to the START, the Short-Term Assessment of Risk, and its variant in scoring approach. Briefly, in the initial version, a continuous 6-point scale was used (i.e., +++ at one pole indicating a considerable strength and ——— at the other pole indicating a substantial risk).

the historical scale because of their relevancy to a high secure hospital setting, and 15 items are divided equally among domains to assess risk for suicide and self-harm, vulnerability, and escape (which includes the possibility of acting in a subversive manner including taking a hostage; Hogue & Allen, 2006). Consistent with the HCR-20, the 45 items are coded as present (yes), absent (no), or partially present (maybe). An overall estimate of risk in the form of a structured judgment ("Tilt High Risk Summary") is made regarding the degree to which an individual presents as high risk across the five areas specified above. Paralleling Version 3 of the HCR-20, the SCJ: Risk also incorporates a risk scenario planning section. The final step in the administration of the SCJ: Risk is to provide clear documentation of risk management plans for each area of high risk. To date, only one empirical evaluation regarding the predictive validity of the SCJ: Risk has been disseminated (Allen & Howells, 2008).

History, Current Behaviour & Future (HKT-30; Werkgroep Pilotstudy Risicotaxatie Forensische Psychiatrie, 2002)

The HKT-30 (see Appendix 1.4) was developed in the Netherlands for use in the judicial context of Terbeschikkingstelling (TBS). TBS is a provision in the Dutch criminal code that allows for a period of treatment following a prison sentence for mentally disordered offenders (van Marle, 2002). The tool comprises 30 items that are scored from 0 to 4 (de Jonge, Nijman, Lammers, & Lucker, 2008). There are 11 static and 19 dynamic factors, mirroring the HCR-20's historical/past, clinical/present, and risk management/future framework. The HKT-30 also provides for the evaluator to render a summary risk rating based on the item scores.

van Marle (2002, p. 89) commented that the HKT-30 is notable in that "it pays a great deal of attention to treatment factors and to factors that arise during the treatment that could have implications for later recidivism." To date, four of the six disseminations that present predictive

validity data for the HKT-30 are not available in English.[8] Three of the investigations were completed by Hildebrand and colleagues (Hildebrand et al., 2005; Hildebrand, Schönberger, & Spreen, 2007; Hildebrand, Spreen, Schönberger, Augustinus, & Hesper, 2006), and the other study was completed by Canton and colleagues (Canton, van der Veer, van Panhuis, Verheul, & van den Brink, 2004).

Structured Assessment of Protective Factors (SAPROF; de Vogel, de Ruiter, Bouman & de Vries Robbé, 2007, as cited in de Vries Robbé et al., 2008)

The SAPROF was developed by clinicians and researchers and contains 17 items, all of which assess protective factors. Its authors envision it as a "positive addition to the HCR-20" (emphasis in original; de Vries Robbé et al., 2008). The Dutch version of the SAPROF was published in 2007, the same year as it was implemented in the van der Hoeven hospital in the Netherlands. An English version is anticipated to be released in October, 2008.

The SAPROF items, which are scored 0, 1, or 2, map onto three scales: Internal factors (5 items); Motivational factors (7); and External factors (5). In addition to scoring individual items, evaluators offer a Final Protection Judgement (low, moderate, high) and an Integrative Final Risk Judgement. The Integrative Judgement is scored using the same three categories and reflects evaluators' judgment of risk based on combined findings offered by the HCR-20 and SAPROF. At least two studies have been completed on the SAPROF, one of which examined predictive validity (de Vries Robbé et al.; van den Broek & de Vries Robbé, 2008).

[8] The two disseminations in English, both of which are conference presentations, are: Hildebrand, Hesper, Spreen, and Nijman (2005); and Schönberger, Hildebrand, Spreen, and Bloem (2008).

Tools for Use with Adults: Specific Forms of Violence

Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997)

The SVR-20 (see Appendix 1.5) is a scheme developed to assess risk for sexual violence among adult sexual offenders. Consistent with the SPJ approach to instrument construction, item selection for the SVR-20 was based on an extensive review of the literature on sexual offenders.[9] The 20 items, which are scored 0, 1, or 2 as is the case for the HCR-20, are grouped into three sections: Psychosocial Adjustment, Sexual Offending, and Future Plans. Whereas items in the Psychosocial Adjustment and Future Plans sections are associated with risk for nonsexual violent and general recidivism, items in the Sexual Offending section are unique to the prediction of sexual violence. Additionally, raters are encouraged to consider so-called critical items, or idiosyncratic factors that may be important for a particular individual's estimate of risk. Finally, consistent with the SPJ model, Boer and co-authors recommended against summing scores on the risk factors, but espoused consideration and integration of the SVR-20 factors and other case-specific factors in making assessments of risk for sexual violence. They further advocated that assessors offer a categorical estimate of risk for sexual violence (i.e., low, moderate, high). Several empirical disseminations that report on the predictive validity of the SVR-20 are available (e.g., Dempster & Hart, 2002; see Hanson & Morton-Bourgon, 2007 for meta-analytic findings).

Risk for Sexual Violence Protocol (RSVP; Hart et al., 2003)

The RSVP (see Appendix 1.6) is a set of guidelines regarding the practice of sexual violence risk assessment. It was developed to assist evaluators in the process of gathering information to make decisions regarding an individual's risk of engaging in sexual violence.

[9] The SVR-20 currently is being revised, with revisions anticipated to be completed by early 2009. The revisions will result in minor changes to content and elaboration of administration procedures (personal communication, S. Hart, August 24, 2008).

Compared to other schemes developed to assess risk for sexual violence, such as the SVR-20, the RSVP includes a stronger focus on risk management. Additionally, development of the RSVP included an expansion in items and content coverage relative to the SVR-20, such that the RSVP comprises additional information regarding non-intimate relationships, social skills, stress, and coercion (Kropp, 2002).

The RSVP is intended to be used with adult men who have a known or suspected history of sexual violence, but can be used with women and adolescents aged 16 and 17 years old (but with caution, given the paucity of research on these groups). The test comprises 22 individual risk factors that map onto five domains: sexual violence history, psychological adjustment, mental disorder, social adjustment, manageability. If appropriate, additional case-specific factors also may be noted. Administration involves six steps: (a) gathering case information; (b) coding the presence of the risk factors, including case-specific risk factors; (c) determining the relevance of the individual risk factors for risk management planning; (d) identifying and detailing the most likely scenarios of future sexual violence; (e) recommending strategies for risk management in light of the information accrued from steps one through four; and (f) documenting one's judgments with respect to overall risk. To date, data from two samples are available regarding the predictive validity of the RSVP (Hart, Jackson, Healey, & Watt, 2008; Kropp et al., 2000, as cited in Kropp, 2002).

Spousal Assault Risk Assessment Guide (SARA; Kropp, Hart, Webster, & Eaves, 1995, 1999)

The SARA (see Appendix 1.7) was developed for use in evaluations of spousal assault. The test's manual defines spousal assault as any actual, attempted, or threatened physical harm perpetrated by a man or woman against someone with whom she or he has, or has had, an intimate, sexual relationship. The SARA comprises 20 risk factors or items that map onto 5 content areas: Criminal History Variables; Psychosocial Adjustment Variables; Spousal Assault;

History Variables; and Alleged (Current) Offence Variables. There is an additional section, Other Considerations, in which the evaluator describes risk factors not included in the SARA but that are relevant for a given case (e.g., presence of stalking behaviour, history of sexual sadism, etc.) and so forth.

The authors of the SARA suggest an assessment procedure based on multiple sources of information and multiple methods of data collection. This is based on the recognition that victims, offenders, and other collateral sources (e.g., children, neighbors) may tend to underreport violence (albeit for different reasons), but that their reports often provide crucial information that is otherwise difficult or impossible to obtain (see Kropp & Gibas, in press).

In completing the SARA, the evaluator renders three kinds of judgments. First, the presence of individual items is coded using the familiar 3-point response format of 0, 1, and 2 according to detailed criteria for each item. Second, the evaluator codes the presence of critical items; these are items that, on their own, are sufficient to compel the evaluator to view the individual's risk of harm as imminent. Critical items are coded using a 2-point format (0 = absent, 1 = present). Finally, summary risk judgments regarding imminent risk of harm posed to the individual's spouse and of imminent risk of harm to some other identifiable person (such as the individual's children) are coded using a 3-point response format (1 = low, 2 = moderate, and 3 = high). Part of the assessment also includes identifying the potential victims if the individual is perceived to be at risk for harming others.

The most recent version of the manual (Kropp et al., 1999) presented an additional step in which the evaluator reviews percentile distributions. Conceptualized as a final cross-check of the evaluator's professional opinion, the summary risk rating for imminent risk for spousal assault is compared to percentile distributions for Total Scores and Number of Factors Present on the SARA. Normative data (see Gibas, Kropp, & Hart, 2008; Kropp & Hart, 2000) are available for inmates (N = 638) and probationers (N = 1,671) who have committed spousal violence. Using the

descriptive cut-offs provided, the evaluator can identify offenders identified as having a large number of risk factors relative to other offenders. The evaluator would offer an explanation for any discrepancy between a percentile ranking and summary risk rating for the given individual. A fair amount of empirical validation has been completed on the SARA (for an overview, see Kropp & Gibas, in press).

Brief Spousal Assessment Form for the Evaluation of Risk (B-SAFER; Kropp, Hart, & Belfrage, 2005)

The B-SAFER is a guide intended to assist professionals working in criminal and civil justice settings to assess and manage risk for spousal assault. The measure, which is an empirically and rationally abbreviated guide based on the SARA, consists of two sections.[10] The Spousal Assault section (5 items) includes factors related to the perpetrator's history of intimate partner violence. The Psychosocial Adjustment section (5 items) consists of factors relevant to psychological and social functioning; these items are associated with risk for violence, broadly defined. Additionally, rare risk factors or those of unique relevance to the case also may be documented, such as access to firearms or sexual sadism. The B-SAFER presents four types of risk management activities to users for their consideration in terms of initiation or implementation: monitoring, treatment, supervision, and victim safety planning.

Administration of the B-SAFER involves a sequence of five steps (Kropp & Hart, 2004). First, background information is documented. Second, the presence of each risk factor is coded for two time periods: within the past four weeks ("Currently") and prior to the past four weeks ("In the past"). Each factor is coded using a 4-level response format (0 = insufficient information available to code the item; "Y/Present; "?"/Unsure, possibly or partially present; and "N"/absent). Third, evaluators estimate the risk to intimate partners if no intervention was taken. Tri-level

[10] A 15 item version being evaluated in Sweden (Belfrage & Strand, 2007) includes a Victim Vulnerability section consisting of the following items: Inconsistent behavior and/or attitude; Extreme fear of perpetrator; Inadequate access to resources; Unsafe living situation; and Personal problems.

categorical ratings (low, moderate, high) are made for estimates of risk for spousal assault within the next two months (imminent risk) and beyond two months (long-term risk). Evaluators also make estimates (low, moderate, or high) of risk for extremely serious assault or death. Finally, risk management strategies are recommended based on the risk level posed and the specific risk factors present, and conclusions are documented. Although research using the B-SAFER is underway (Belfrage & Strand, 2007), limited data regarding predictive validity of the instrument are available (Kropp, 2003).

Guidelines for Stalking Assessment and Management (SAM; Kropp, Hart, & Lyon, 2008)

The SAM assists professionals from criminal justice, mental health, and security disciplines in evaluating risk for stalking by offering a "systematic, standardized, and practical framework for gathering and considering information when making decisions about stalking risk" (Kropp et al., 2008, p. 5). The scheme comprises 30 risk factors that are divided equally into 3 domains. The first domain, Nature of Stalking, assists the evaluator in characterizing the seriousness of the alleged perpetrator's stalking behaviour, and includes items related to the pattern of behaviour evident in the index offence. The second domain, Perpetrator Risk, examines characteristics of the perpetrator potentially associated with decisions to engage in stalking; items in this domain address the psychosocial adjustment and background of the perpetrator. Finally, items in the Victim Vulnerability Factors domain reflect the psychosocial adjustment and background of the victim and are included in order to identify the victim's unique circumstances and vulnerabilities as well as those characteristics that may be associated with decisions to engage in self-protective behaviour. The SAM additionally includes a section where evaluators can document factors not included in the scheme but that are relevant to the given case.

The scheme is intended for use with men and women of any sexual orientation who have a known or suspected history of stalking. Such individuals typically should be at least 18 years of

age, although the SAM may be used when the primary perpetrator is an adolescent (though research on juvenile perpetrators is limited). In cases that involve multiple primary perpetrators and/or multiple primary victims, the SAM should be completed separately for each. When the evaluation involves corporate victims, Kropp et al. (2008) suggest that victims may need to be grouped based on a relevant characteristic, such as working in the same office, with separate SAMs to be completed for each group.

In terms of rating the items, the presence of the 30 risk factors is coded using a 3-point format ("Y"/definitely or conclusively present; "?"/possibly or partially present; "N"/ no evidence that the risk factor was present). Additionally, the first two domains - Nature of Stalking and Perpetrator Risk Factors—are coded "Current" versus "Previous" to reflect possible changes over time in stalking behaviour and risk factors. More specifically, "Current" pertains to risk factors during the most recent pattern of stalking behaviour, whereas "Previous" ratings reflects the status of the risk factors during past stalking of any victim. Only "Current" ratings are made for items in the Victim Vulnerability Factors domain.

Administration of the SAM involves five steps: (a1) gathering and documenting basic case information, (b) coding the presence of the 30 factors and any additional case specific risk factors, (c) identifying and detailing the most likely scenarios of future stalking, (d) presenting recommendations for risk management strategies on the basis of the information documented during the preceding steps, and (e) rendering judgments regarding overall risk (low or routine; moderate or elevated; high or urgent). This relatively new tool has already been examined empirically (e.g., Belfrage & Strand, in press) but, to date, findings regarding predictive validity have not been published.

Workplace Risk Assessment-20 (WRA-20; Bloom, Eisen, Pollock & Webster, 2000)

The authors of the WRA-20 describe the guide as one that "quantifies an organization's risk for a violent occurrence on its premises" (http://workplace.calm.to/products/wra20.html). Information available on the website further indicates that the measure "was designed for use by human resources professionals, managers, security, corporate health, union officials and workplace consultants...It is intended to foster creative and frank discussion amongst the various stakeholders, all of whom have an interest in belonging to and participating in safe organizations." The authors also indicate that WRA-20 "provides a motivational framework within which the key stakeholders in the workplace violence program can contribute their wisdom and experience towards avoiding a potentially serious and even life-threatening event." To date, no data are available regarding its reliability or validity.

Employee Risk Assessment-20 (ERA-20; Bloom, Webster, & Eisen, 2002)

The ERA-20 is described by its authors as the "other side" of the WRA-20, "designed specifically to evaluate an individual worker's risk for engaging in violence in the workplace" (http://workplace.calm.to/products/era20.html). The ERA-20 consists of 20 items:

> most relevant to evaluating an individual's risk for workplace violence. The rationale for including an item is drawn from the scientific literature currently amassing in the fields of risk assessment and violence prediction, from the developing field of workplace violence risk assessment, and from a wealth of clinical experience. (¶ #)

It is intended to be used by "trained clinicians who in their consultative work make considered judgements about workplace risk" as well as to assist "human resources, corporate health, corporate security, union officials, and employment lawyers to get a broad sense of how much concern an organization is justified in having about a worker's behaviour." As is the case for the WRA-20, empirical evaluation of this scheme has not been completed.

Tools for Use with Adolescents

Whereas several SPJ tools have been developed to assess risk for various types of violence perpetrated by adults, only two tools (reviewed below) exist for use with adolescents. One tool was developed for assessing risk for general violence, and the other, for sexual violence.

Structured Assessment of Violence Risk in Youth (SAVRY, Borum et al., 2003)

SAVRY "is designed to assist professional evaluators in assessing, and making judgments about, a juvenile's risk for violence" (Borum et al., p. 6). The SAVRY emphasizes dynamic risk/needs factors in recognition of the significant developmental changes that occur in physical, intellectual, social, and emotional domains during adolescence (see Appendix 1.8).

The scheme includes 24 risk factors or items grouped into three domains: Historical (10 items), Social/Contextual (6 items), Individual/Clinical (8 items). Additionally, the SAVRY includes the Protective Factors domain, which consists of 6 factors associated with reductions of risk when present for a given adolescent. Items on the Historical, Social/Contextual, and Individual/Clinical domains are coded using a three-level format (high, moderate, low). Items on the Protective Factors domain are coded dichotomously as present or absent. Evaluators also are invited to list additional risk factors that are relevant for a given individual. Finally, evaluators make a summary risk rating (low, moderate, high) based on all risk and protective factors. To date, over 30 papers have been disseminated in which the SAVRY was discussed and/or evaluated empirically.

Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR; Worling & Curwen, 2001)

The ERASOR (see Appendix 1.9) was developed to assist evaluators in assessing risk of sexual recidivism. It may be used only with individuals between the ages of 12 and 18, and only with those who already have committed a sexual assault. The most recent version of the manual

indicates that the ERASOR was "developed in a similar fashion to the SVR-20" (Worling & Curwen, p. 4). The ERASOR comprises 25 risk factors that map onto five categories: (a) Sexual Interests, Attitudes, and Behaviours; (b) Historical Sexual Assaults; (c) Psychosocial Functioning; (d) Family/Environmental Functioning; and (e) Treatment. Additionally, evaluators may record and consider case-specific risk factors.

Individual items are coded using a 4-level format: Present, Partially/Possibly Present, Not Present, and Unknown. Summary risk ratings also are made using the low/moderate/high categorization. The authors note that although "it is anticipated that there will be a general relationship between the number of high-risk factors and the rating of risk such that more high-risk indicators suggest higher risk... the final decision will be more dependent on the combination of risk factors rather than just the number" (Worling & Curwen, 2001, p. 5). The authors, following Boer et al. (1997), also note the possibility that the presence of a single, especially salient risk factor could be indicative of high risk. The body of findings regarding the empirical validation of the ERASOR is small but developing (e.g., McCoy, 2007; Morton, 2003).

Tools for Use with Children

Early Assessment Risk List for Boys (EARL-20B; Augimeri, Koegl, Webster, & Levene, 2001)

The EARL-20B (see Appendix 1.10) is a decision aid developed to assist evaluators in considering relevant risk and needs factors when evaluating risk for antisocial and aggressive behaviour among 6- to 12-year-old boys who are exhibiting disruptive behaviour problems, and in developing effective clinical risk management plans for the boys and their families (see Augmeri, Enebrink, & Walsh, in press, for an overview of the development and testing of early versions of the EARL-20B, as well as a summary of findings regarding the scheme's reliability and validity). The EARL-20B consists of risk and needs factors organized according to their relevance to the Family (6 items), the Child (12 items), and Amenability (2 items).

Items in the Family domain assess the extent to which the boy has received effective nurturance, support, supervision, and encouragement. Also pertinent to assessment with this domain is a consideration of the amount of stress with which the family is coping, level of support available to manage stressors, and the extent of antisocial attitudes and behaviours evident among family members. Items in the Child domain address individual risk factors, which include contextual risk factors such as quality of the boy's neighborhood. Finally, the two items that comprise the Amenability domain assess the anticipated responsivity of the boy and his family; both their ability and willingness to engage in treatment and to benefit from planned interventions is considered.

Similar to other SPJ tools, items are rated on a 3-point scale (0 = not present, 1 = partially present, 2 = definitely present). Additionally, any factor can be identified as one that indicates "Critical Risk." The authors included this feature on the EARL-20B to afford "the opportunity to adapt assessment and treatment to meet the requirements of a particular case, a feature viewed as essential to evidence-based clinical practice (American Psychological Association, 2006)" (Augmeri et al., in press, p. 6). The authors recommend that evaluators using the EARL-20B provide an estimate of the boy's overall level of risk for behaving in an antisocial manner (low, moderate, high) that is based on the nature and frequency of item endorsements, the particular risk/needs factor pattern, and case-specific factors. The EARL-20B can be completed in approximately 15 to 30 minutes when the rater(s) has good knowledge of both the child and the risk tool (Hrynkiw-Augimeri, 2005).

Although scores can be summed to yield a total score with a range of 0-40, the authors advise that "the 20-item scale should not be used in a mechanical way to determine the availability or intensity of treatment (i.e., by using the total score as the determinant)" (Hrynkiw-Augimeri, 2005, p. 43). Rather, an estimate of the child's overall level of risk (low, moderate, high) should be offered and used for clinical decision-making purposes. Commentaries and

descriptive data about the EARL-21B are available (Augimeri, Enebrink, & Walsh, in preparation; Augimeri, Jiang, Koegl, & Carey, 2006; Augimeri, Koegl, Ferrante, & Slater, 2006; Augimeri, Koegl, Levene, & Webster, 2005; Enebrink, Långström, & Gumpert, 2006; Enebrink, Långström, Gumpert, & Hulten, 2003; Enebrink, Långström, Hulten, & Gumpert, 2006; Hrynkiw-Augimeri, 1998; Koegl, Webster, Michel, & Augimeri, 2000), and empirical findings regarding predictive validity (Enebrink, Långström, Neij, Grann, & Gumpert, 2006; Hrynkiw-Augimeri, 2005) are accruing.

Early Assessment Risk List for Girls, Version 1 Consultation Edition (EARL-21G; Levene, Augimeri, Pepler, Walsh, Webster & Koegl, 2001)

Developed in concert with the EARL-20B, the EARL-21G is a parallel assessment tool for girls. All but two of the item headings are the same as the EARL-20B headings. However, item content and coding differ between the two schemes, reflecting their gender sensitive nature (to the extent that the basis for doing so is offered by the scientific literature). The two items unique to the EARL-21G are Caregiver-Daughter Interaction and Sexual Development. The other deviation in terms of item headings is that the item that appears as 'Authority Contact' on the EARL-20B was subsumed under 'Antisocial Behaviour' on the EARL-21G. At this point, only a few disseminations that focus on the EARL-21G are available (Yuille, 2008; Levene et al., 2001; Levene, Walsh, Augimeri, & Pepler, 2004).

Non-SPJ Risk Assessment Tools

Following the definition of SPJ tools used in the present project, non-SPJ tools consist of any measures developed to assess risk for violence to others that were not developed according to the principles of the SPJ approach outlined earlier. Given the strict focus of the present project on tools that were developed explicitly within the framework of the SPJ model to assess risk for

violence to others, many useful risk assessment tools consequently were not included.[11] For example, many non-SPJ risk tools that nevertheless are consistent with the SPJ approach in that they include dynamic risk factors were excluded from the present review. Just a few examples of such measures include the Dynamic Appraisal of Situational Aggression (DASA; Ogloff & Daffern, 2006), Dynamic Risk Assessment and Management System (DRAMS; Lindsay et al., 2004); Clinical Inventory of Dynamic Reoffending Risk Indicators (CIDRRI, Philipse, Koeter, van den Brink, & van der Staak, 2004); and the Structured Assessment of Risk and Needs[12] (SARN; Thornton, 2002).

Two actuarial tools intended to be used for decision-making in conjunction with clinical discretion are the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995), as well as related tools in the LSI family, and the Classification of Violence Risk (COVR; Monahan, Steadman, Appelbaum et al., 2005). The LSI-R yields a total score used to classify individuals into levels of risk associated with projected probabilities of violence. However, users of this tool are allowed to exercise clinical discretion at the final step of the evaluation by applying a clinical over-ride should they deem it appropriate.

The COVR is another actuarial tool whose application is not disjoint from clinical judgment. The development of the COVR (Monahan, Steadman, & Appelbaum, 2001; Monahan, Steadman, Silver, et al., 2005a; Monahan, Steadman, Robbins, et al., 2005b) represents the most sophisticated and comprehensive effort to date to create an actuarial tool. Development of this tool, which used data from adults released from psychiatric hospitalization to the community, relied on classification tree analysis that allows multiple variables to be considered concurrently. .

[11] As detailed in the Methods section below, although non-SPJ tools were excluded from the literature search undertaken to identify studies to code for the present meta-analysis, non-SPJ tools *were* coded if predictive validity data for them were presented in the same study selected for the meta-analysis on the basis that data on a SPJ tool were presented. Non-SPJ data of this sort were included for comparative purposes because direct comparisons of measures within the same sample yield the cleanest comparison possible between performance of SPJ and non-SPJ tools.

[12] Originally named Structured Risk Assessment instrument (S. D. Webster et al., 2006).

The COVR produces an actuarial estimate of risk, but is intended to be just one piece of information upon which clinicians base their decisions (i.e., it is used to inform clinical judgment).

Additionally, there are several risk assessment schemes that may be described as guided clinical judgment assessment tools because clinical judgment was used to select and weight the risk factors contained in the scheme, but which would not be conceptualized as SPJ tools. An important distinction between such schemes and those adhering to the SPJ approach is that, in these non-SPJ schemes, the person completing the risk assessment uses rules rather than clinical judgment to render an estimate or prediction of the level of risk. These explicit rules for combining items are determined when the scale is constructed, at which time item scores or relative weights are assigned (though based on clinical judgment) during the scale development. The risk evaluator scores each item according to fixed criteria, and then sums item scores to yield a total score that corresponds to a predetermined category or level of risk.

Examples of risk schemes of this genre include the Registrant Risk Assessment Scale (RRAS; Ferguson, Eidelson, & Witt, 1998), the Sex Offender Screening Tool (SOST; Epperson, Kaul, & Huot, 1995), and the Vermont Assessment of Sex Offender Risk (VASOR; McGrath & Hoke, 2001). For example, a panel of mental health experts assigned item weights and risk category cut-off scores for the VASOR.[13] Another measure described by its authors as a guided clinical scheme is the Multifactorial Assessment of Sex Offender Risk for Recidivism (MASORR; Barbaree & Seto, 1998; Barbaree, Seto, Langton, & Peacock, 2001), which incorporates a statistically generated probability of recidivism based on actuarial risk assessment tools.

[13] Decisions were based on the panel's determination of which offenders among the developmental sample of 122 sex offenders should be released to the community or incarcerated (McGrath & Hoke, 2001). Risk category classification was a function of rank ordered mean VASOR scores.

Current State of the SPJ Literature: Assessing and Aggregating Accumulated Knowledge

Given the tremendous productivity of researchers investigating the SPJ approach to violence risk assessment during the past several years, as well as the wide-spread implementation of the approach into clinical practice (Crocker et al., n.d.; Desmarais, Nicholls, & Brink, 2007; Müller-Isberner & Jockel, 1997; RMA, 2007; Webster, Nicholls, et al., 2006), a synthesis of the empirical support for this approach would be valuable. Whereas qualitative reviews of the literature can be worthwhile in summarizing available findings, a more detailed and methodologically rigorous approach is quantitative analysis (see, e.g., Lipsey & Wilson, 2001; Rosenthal, 1991; Rosenthal & DiMatteo, 2001).

According to the *Dictionary of Epidemiology*, meta-analysis is defined as "the statistical synthesis of the data from separate but similar, i.e., comparable studies, leading to quantitative summary of the pooled results" (Last, 2001, p. 114). Glass (1976a, 1976b) introduced the term *meta-analysis* approximately 30 years ago during an address in which he highlighted the need for improved synthesis of research results (see Chalmers, Hedges, & Cooper, 2002). Most contemporary statistical techniques have their origins in a dissemination of Gauss' and Laplace's work that was published in 1861 by the British Astronomer Royal (Airy, 1861). The basic procedure of meta-analysis involves taking the results from independent studies, transforming them into a common metric or statistic (an effect size[14]), and then comparing the results across studies (see Rosenthal & DiMatteo, 2001).

Potentially Moderating Variables

Although establishing overall rates regarding the predictive validity of the SPJ decision-making model is useful, it is equally important to know whether, and, if so, how, the predictive

[14] Effect sizes are estimates of the magnitude of the association between two variables.

validity of the model and specific SPJ tools function over a variety of contexts and under a range of circumstances. Investigations regarding which factors may influence predictive accuracy in this fashion centre on the identification of moderator variables. Several such variables would be worthwhile to study for their potential to inform clinical practice (i.e., how and when SPJ tools should be applied in real-world contexts) and science (i.e., how research on the use of SPJ tools may be improved). Previous research findings, as well as the likely benefit of accruing knowledge that would inform clinical and research practices related to violence risk assessment, guided the choice of the following moderator variables.

Factors Affecting Clinical Practice

Results of several studies indicate that *gender* is an important variable to study within violence risk assessment because clinicians have been found to underestimate the risk of future violence in female psychiatric patients (Coontz, Lidz, & Mulvey, 1994; Elbogen, Williams, Kim, Tomkins, & Scalora, 2001; McNeil & Binder, 1995; Skeem, Schubert, et al., 2005; see also Odgers, Schmidt, & Reppucci, 2004; Teasdale, Silver, & Monahan, 2006; Strand & Belfrage, 2001; Nicholls, 2001; Nicholls, Ogloff, & Douglas, 2004). Researchers have noted that although there does seem to be a genuinely lower rate of violence among women in society as a whole, this pattern does not seem to hold for psychiatric populations (Nicholls, 2001; Robbins, Monahan, & Silver, 2003), and it is hypothesized that evaluators may adjust their predictions based on the former but not the latter.

Another explanation for mental health professionals' tendency to underestimate women's violence is that violent women are more likely to be violent with family members in the home, and therefore their offending is less observable (Hiday, Swartz, Swanson, Borum, & Wagner, 1998; Robbins et al., 2003). The importance of examining predictive accuracy as a function of gender is especially noteworthy for actuarial measures. For example, the authors of the VRAG,

which was calibrated using a sample of men, reported that the measure was unrelated to predictive accuracy for violent recidivism among female offenders (Harris et al., 2002).

Although there are more circumstances under which risk assessments potentially may be required for adults, risk assessments also are performed with regularity for adolescents (e.g., release decision-making, transfer to adult court). Although the impact of *age* on predictive accuracy has been examined with some actuarial tools (e.g., Hanson, 2006), no such investigations have been conducted within the SPJ realm. As such, whether the sample under study consisted of adult, adolescents, or children will be examined. The *country* in which the data were collected also seems to warrant empirical investigation. Most contemporary risk assessment technologies were developed in North America (with almost all SPJ tools having been developed in Canada), which gives rise to concerns regarding test generalizability. Although not developed for use as a risk assessment tool, results of a meta-analysis of the predictive accuracy of the Psychopathy Checklist-Revised (PCL-R; Hare, 2003) and its Short Version (PCL:SV; Hart, Cox, & Hare, 1995) as a risk factor for institutional violence nevertheless are relevant to this concern. Guy, Edens, Anthony, and Douglas (2005) found that the predictive utility of PCL-based measures for violent infractions was relatively weaker within US prisons (r_w = .11), and comparably stronger in non-US prisons (r_w = .23).

Again considering the psychopathy literature, several studies provided evidence of cross-national differences in the assessment of this personality disorder. Cooke and Michie (1999) observed a lower prevalence of PCL-R defined psychopathy among Scottish prisoners than American prisoners and forensic patients, prompting a recommendation that cut scores be reduced by approximately five points for Scottish examinees. Additional research by Cooke and colleagues demonstrated that participants in the U.K., on average, obtained lower total scores for the same level of the underlying trait compared with participants in North America (Cooke & Michie, 1999; Cooke, Michie, Hart, & Clark, 2005). Because many SPJ risk tools have been

translated[15] and used in numerous countries outside of North America,[16] studying whether the measures perform comparably across nations is important.

The nature of the *clinical setting* (e.g., forensic psychiatric, civil psychiatric, correctional) from which participants were drawn will be studied. Although a meta-analysis (Bonta et al., 1998) found that the major predictors of recidivism were the same for mentally disordered and nondisordered offenders, it does not necessarily follow that the same method of risk assessment would yield equally useful degrees of predictive validity in different contexts with different types of people. As just one example, although past violence is a well-established risk factor for future violence, historical data often are not available in crisis settings (such as emergency rooms), which conceivably would affect the risk assessment process under such circumstances (see Elbogen, Huss, Tomkins, & Scalora, 2005; Wooten et al., 2008). In addition, different types of settings where violence occurs would be expected to vary in terms of important contextual factors that may affect the outcome of violence, such as security level in institutional settings (Guy et al., 2005). As such, *violence location* will be examined to investigate there is a moderating effect in terms of whether violence occurs in the community or an institutional setting.

Finally, whereas many actuarial tools can be completed solely on the basis of file information given their item content, SPJ tools include items whose scoring presumably would benefit from face-to-face evaluation of the individual (e.g., active symptoms of mental illness). As such, the *sources of information* (file only, file and interview) used in the assessment will be evaluated.

[15] For example, there exist at least 16 authorized translations of the HCR-20, and others are in progress (personal communication, K. Douglas, August 23, 2008; see http://kdouglas.wordpress.com/hcr-20).

[16] Some examples include: Argentina (Folino, Marengo, Marchiano, & Ascazibar, 2004), Belgium (Claix, Pham, & Willocq, 2002), Germany (Müller-Isberner & Jockel, 1997), Norway (Urheim, Jakobsen, & Rasmussen, 2003), Sweden (Belfrage, Fransson, & Strand, 2000; Belfrage, 1998; Belfrage & Douglas, 2002; Dernevik, 1998; Dernevik, Grann, & Johansson, 2002; Douglas, Strand, Belfrage, Fransson, & Levander, 2005; Grann, Belfrage, & Tengstrom, 2000; Strand & Belfrage, 2001), the Netherlands (de Vogel & de Ruiter, 2004, 2005; de Vogel, de Ruiter, Hildebrand, Bos, & van de Ven, 2004; de Vogel, de Ruiter, van Beek, & Mead, 2004; Hildebrand, de Ruiter, & de Vogel, 2004; Hildebrand, de Ruiter, & Nijman, 2004), and the U.K. (Cooke & Michie, 1998).

Factors Affecting Research

Although likely confounded at least in part by the type of information used to complete the risk assessment, *study design* (retrospective, pseudo-prospective, true prospective) also will be investigated for its potential influence on estimates of predictive validity. In a true prospective design, the risk assessment is completed based on current information known at some time, T, and the criterion occurs at some point, T+1, in the future. In a pseudo-prospective design, the risk assessment is completed at some time, T, based on information available at a point in the past, T-1, and the criterion is coded at a point subsequent to T-1. This design is intended to mimic a truly prospective design, in that: (a) the criterion of interest occurs after the point in time (T-1) in which the information that forms the basis of the risk assessment coding existed, and (b) researchers are (typically, ideally) blind to outcome. In retrospective (also called postdictive) studies, the criterion occurs at a point prior to the period on which the risk assessment is based. This is the least strong design, because the 'outcome' is already known, and happened in the past, which does not represent the manner in which risk assessment instruments are intended to be used in practice.

Finally, *allegiance* (defined as being present when a study is conducted by the people who developed the instrument) will be examined. The dramatic impact that allegiance can have on reported empirical findings was studied first in the psychotherapy outcome literature. Luborsky et al. (1999) found that researchers' therapy allegiance (operationalized using reprint ratings, self-ratings, and ratings by colleagues) was associated strongly with outcomes reported in studies in which the efficacy of different psychosocial treatments was compared, with allegiance accounting for 69% of the variance in differences between treatments.

Recently, the impact of allegiance has been investigated within the violence risk assessment literature. Using meta-analysis to examine the predictive validity of the Violence Risk Appraisal Guide (VRAG; Harris et al., 1993), the Sex Offender Risk Appraisal Guide (SORAG;

Quinsey, Harris, Rice, & Cormier, 1998), and the Static-99 (Hanson & Thornton, 1999), Blair, Marcus, and Boccaccini (in press) reported that effect sizes were significantly larger in studies conducted by the instruments' authors than in studies conducted by independent researchers (*r*s of .37 versus .28, respectively). Moreover, no other design or sample characteristics significantly moderated the relation between scores and recidivism. This dissertation will examine whether the allegiance effect observed by Blair et al. (in press) for actuarial measures also may be present in research on SPJ tools.

The Present Research

As outlined above, the benefits of focusing on risk prevention, management, and reduction are clear. However, a prerequisite for such efforts is the accurate identification of risk status. To that end, the primary purpose of this dissertation is to investigate the performance, in terms of predictive validity, of an approach to violence risk assessment that views the tasks of risk prevention, management, and reduction as central to and intertwined with the prediction of violence risk—i.e., the structured professional judgment approach.

Meta-analysis will be used to examine the evidence regarding predictive validity of all known SPJ tools, the first such study of this kind. The present research will improve on previous meta-analyses that included SPJ measures by its comprehensive review of the literature and inclusion of studies irrespective of the age, gender, or psychiatric characteristics of the sample. This project also will go beyond previous meta-analyses that examined prediction approaches to violence risk assessment by investigating whether there may be certain variables that systematically affect the predictive validity of the SPJ approach. In sum, this investigation will examine the following five research questions:

1. What is the aggregate estimate of the magnitude of predictive accuracy for (a) the SPJ model (b) judgments based on summary risk ratings of SPJ tools and (b) judgments based on numeric scores of SPJ tools?

2. Do any of the following variables moderate the predictive validity of the SPJ model, and, if so, in what way?
   a. Gender
   b. Age group
   c. Nationality/Country of data collection
   d. Clinical population type
   e. Violence location
   f. Source of information used to complete risk tool
   g. Study design
   h. Allegiance
3. What is the estimate of the magnitude of predictive accuracy for each SPJ tool for various types of antisocial outcomes?
4. Is the performance of "broad" and "specific" SPJ tools maximized when the type of violence for which the tool was developed to assess parallels the type of outcome under consideration?
5. How does the overall estimate of the magnitude of predictive accuracy for SPJ tools compare to that for non-SPJ tools used in the same study?

METHODOLOGY

Eligibility Criteria

Studies were coded for the meta-analysis if they met two inclusion criteria. First, they needed to present quantitative data on the association between a score or judgment rendered using an SPJ tool (as defined above) for assessing risk for engaging in any type of antisocial behaviour. Second, an effect size needed either to have been reported directly or sufficient information had to have been presented to allow an effect size to be estimated.

Search Procedure

Published Literature

Published disseminations were located in two primary ways. First, searches were conducted of the following nine computerized literature databases for studies published by March 1, 2008: PsycINFO (covers scholarly publications in the behavioural and social sciences), MEDLINE (covers biomedical and health literature; an overview of all databases provided by the U.S. National Library of Medicine also is available); National Criminal Justice Reference Service (covers all aspects of law enforcement, crime prevention and security, criminal justice, and juvenile justice); Criminal Justice Abstracts (index to all aspects of criminal justice); Sociological Abstracts (theoretical and applied sociology, social science, and policy science); Health Source: Nursing/Academic Edition (covers nursing and allied health topics); SocIndex (indexes articles, books, and conference proceedings in Sociology and related fields); Mental Measurements Yearbook (descriptive and evaluative information about tests); and Web of Science (Science, Social Sciences and Arts and Humanities Citation Indexes). The stems of the following identifier

and subject words were used in separate and combined searches: Predict*; Antisocial; Sex*; Offend*; Aggress*; Violen*; Risk; Danger*; Recidiv*; Rearrest; Arrest; Stalk*; Offend*; Crim*; Prison*; Convict*; Bully*; child abus*; workplace. In addition, separate searches were conducted using the full name and acronym or abbreviation of specific measures identified via the above procedures. Second, reference lists of primary studies included in the meta-analysis and peer-reviewed narrative review articles were examined for relevant studies.

To explore whether it would be beneficial to use an additional search strategy that involved examining tables of contents of relevant journals, a pilot search was completed. The tables of contents of all issues of seven journals published during 2004 – 2006 (inclusive) were examined. The specific journals (*American Journal of Psychiatry*, *Behavioral Sciences and the Law*, *British Journal of Psychiatry, Criminal Justice and Behavior*, *Journal of Abnormal Psychology, Journal of the American Academy of Psychiatry and Law,* and *Law and Human Behavior*) were selected because of their differing target audiences (e.g., psychiatrists, psychologists, general clinical practitioners, specialized forensic mental health professionals, etc.) and to capture journals that covered general topics in clincial psycholgy and psychiatry, as well as more specialized journals that focused on topics more relevant to law and psychology. The search yielded titles of several relevant studies. Without exception, all studies identified via the table of contents search that met inclusion criteria already had been identified during the previous search steps detailed above. Because this strategy did not identify any new studies that met inclusion criteria, it was not pursued.

Unpublished Literature

Four search strategies were used to identify unpublished disseminations. First, the annotated bibliography of research maintained by Douglas and colleagues (2007) on the HCR-20

that contains summaries of unpublished research was examined. Second, programs of relevant conferences during 2001-2005[17] were examined for presentations that had not subsequently appeared in peer-reviewed publications. Programs of the following organizations were downloaded from their website or, when not available, were requested from a representative of the organization: American Psychology-Law Society; American Psychological Association; Canadian Psychological Association; American Academy of Forensic Psychology; American Academy of Forensic Sciences; Australian and New Zealand Association of Psychiatry, Psychology, and Law; European Association of Psychology and Law; International Association of Forensic Mental Health Services; International Academy of Law and Mental Health; Society for the Scientific Study of Psychopathy; American Psychological Society; American Psychiatric Association; and Canadian Psychiatric Association.

Third, a request for studies was sent to two groups of forensic mental health professionals. One solicitation was sent to the list serve of the American Psychology-Law Society; the second was sent on my behalf by the president of the International Association of Forensic Mental Health Services to a group of professionals whose research and/or clinical interest include violence risk assessment. In both instances, professionals were invited to e-mail me if they had unpublished data or findings (conference presentations, manuscripts in progress or that were never published) that could be included in the meta-analysis.

Fourth, four databases or online catalogues were searched to identify unpublished dissertations. The 'Digital Dissertation' database was searched for projects completed in North America. The Universal Index of Doctoral Dissertations in Progress (http://www.phddata.org) is described as "a database of doctoral dissertations in progress around the world." This site is not a

[17] This time frame represented the 5 year period immediately preceding the time when the initial search for fugitive literature for the present project was completed. The exception to this rule was that relevant studies listed in programs for two conferences held in 2008 (International Association of Forensic Mental Health Services and American Psychology-Law Society) were included as well.

comprehensive listing of such material, as it only contains dissertations that have been registered by their authors. The British Thesis Service (http://www.bl.uk/services/document/brittheses.html) is described as holding "full text of more than 170,000 doctoral theses, mainly from the 1970s to the present day. Almost all UK universities make their theses available through the Service." Finally, catalogues of academic or national libraries in Australia, New Zealand, and the European Union were searched using Libweb (http://lists.webjunction.org/libweb/), which lists "over 7400 pages from libraries in over 125 countries." Boolean searches were conducted using the same key words as specified above.

Coding Rules and Procedures

Predictor Variables

When available, effect sizes were coded for risk assessment measures' total scores, scale scores, and summary risk ratings. In addition to coding as many effect sizes as possible for the particular SPJ tool(s) under investigation in a given study, multiple effect sizes for non-SPJ risk assessment tools also were calculated or estimated from each study when presented. For example, if a study was included in the meta-analysis because it investigated the SAVRY, but it also investigated an actuarial risk tool, multiple effect sizes were included for both risk assessment measures. Additionally, for comparison purposes, effect sizes for the PCL family of measures and unstructured clinical judgment also were calculated or estimated when available.

Criterion Variables

A seven category coding scheme was developed subsequent to identification of studies that met inclusion criteria based on familiarity with the types of outcomes typically reported in the literature. Most research studies in the violence risk assessment literature reported a global outcome type that served as an indicator of any form of antisocial behaviour in the community or

an institution. Therefore, the most general level of the hierarchy comprised a total or 'any' outcome category. For studies that presented effect sizes for both non-mutually exclusive and mutually exclusive criterion variables (e.g., sexual violence, non-sexual physical violence, and any violence), the most inclusive category (in this example, 'any violence') was coded at the general level of the hierarchy. For studies that only presented effect sizes for mutually exclusive categories (e.g., sexual violence and non-sexual physical violence) the average of the effect sizes of all categories was coded at the general level of the hierarchy.

Four of the criterion categories were developed to reflect non-redundant, mutually exclusive types of antisocial behaviour. Effect sizes were coded into the 'physical violence, including sexual violence' category if the outcome variable was described as representing physically aggressive behaviour toward another person; if it was made explicit that sexually aggressive acts were included among acts of physically aggressive behaviour; or if it was not specified whether sexually aggressive acts were included among acts of physically aggressive behaviour. Effect sizes were coded into the 'physical violence, excluding sexual violence' category if the outcome variable was described as representing physically aggressive behaviour toward another person and it was made explicit that acts of sexual violence were not included in the operationalization of violence. The 'sexual violence' category represented any act of sexually aggressive behaviour. The 'non-violent' category included acts that studies described as verbal threats, verbal abuse, verbal aggression, property destruction, and any crimes of a non-physically violent nature (e.g., theft, fraud, drug offenses).

The remaining two mutually exclusive categories in the coding schemes were 'intimate partner violence' (IPV) and 'violence.' Effects were coded into the IPV category if a study described the criterion as any type of aggression perpetrated against an intimate partner. This category of behaviour may include any form of aggression (e.g., physical aggressive violence, verbal threats, etc.). Finally, the 'violence' category was created to code effect sizes from studies

that described the criterion as being operationalized using the definition of violence presented in the HCR-20 manual ("actual, attempted, or threatened harm to a person or persons," Webster et al., 1997, p. 24). Although the manual's definition is intentionally broad, many studies indicated that the actual types of outcome behaviours observed were of a more serious nature (e.g., physically aggressive violence). Rather than code effect sizes for this type of operationalization only at the general level of 'any' category, and thereby abandon a certain amount of specificity, the separate 'violence' category was created.

Potentially Moderating Variables

As described in the introduction, a host of variables were identified *a priori* to be examined for their potential to moderate predictive accuracy. Whenever possible, multiple effect sizes were calculated or estimated from a single study to allow for examination of certain characteristics as they may relate to the strength of the association between risk assessment and violence (see below). All of the variables listed below were chosen for inclusion because they might, on conceptual, empirical (i.e., past literature), or methodological grounds, impact accuracy. Two major categories of variables were coded: variables affecting clinical practice, and those more directly relevant to research practice.

Factors Affecting Clinical Practice

Five variables that are directly implicated in the use and application of risk assessment tools in clinical practice were analyzed. *Gender* was coded as samples that consisted of all women, all men, or both men and women. *Age* was coded as to whether samples comprised adults, adolescents, or children. The *country* in which the data were collected was coded dichotomously as Europe or North America. The nature of the *clinical setting* comprised four levels: forensic psychiatric, civil psychiatric, correctional, and mixed/other/not specified. *Violence location* was coded to reflect violence that occurred in the community or in an

institution (i.e., hospital, jail, or prison). Finally, the *sources of information* used to complete the risk assessment were coded as file only or both file and interview.

Factors Affecting Research Practice

Two variables related to the way in which research investigations are carried out were coded. *Study design* comprised three levels: retrospective, pseudo-prospective, and true prospective. Finally, whether any of the study's authors also was an author of the risk assessment measure under investigation was coded as *allegiance*. Because numerous official translations of manuals of many SPJ tools are available, another level of the allegiance variable examined whether any of the study's co-authors also was an author of the translated version of the risk assessment tool.

Interrater Reliability

A coding booklet (see Appendix 1.11) was written that incorporated the domains of information detailed above. All studies were coded by the author. Interrater reliability (IRR) analyses were based on ratings for 24 studies (21% of disseminations included in the meta-analysis) completed by raters blind to the author's ratings. Half (n = 12) of the IRR cases was selected randomly from the disseminations that focused on the HCR-20, and the other half of IRR cases was selected randomly from the disseminations that investigated other SPJ tools. For the HCR-20 studies, the author's ratings were compared to consensus ratings of two advanced graduate students in forensic clinical psychology (K. Reeves and N. Nikolova). Two practice cases were completed prior to the 12 IRR cases. For the 12 other SPJ studies used for IRR, the author's ratings were compared to ratings of a researcher with a Master's degree in forensic psychology (N. Cook). Eight practice cases were completed prior to these 12 IRR cases. Fewer practice cases were completed for the HCR-20 studies because the three raters had previous

experience completing IRR ratings with one another, and two practice cases were deemed to be sufficient for the present project.

Consistent with approaches followed by other meta-analysts who have investigated the predictive validity of forensically relevant measures (e.g., Edens, Campbell, & Weir, 2007; Campbell et al., 2007), IRR was evaluated by examining the overall rate of agreement. Thirty-seven variables were evaluated for agreement across 24 studies, thereby yielding 888 observations (444 each for the HCR-20 and other SPJ studies). Consensus ratings could not be completed for type of criterion category for the HCR-20 studies because the IRR raters followed a different coding scheme for that single variable,[18] agreement regarding operationalization of violence was evaluated, however.

Because an index of the frequencies of raw agreement is conceptually easy to understand, Yeaton and Wortman's (1993) formula[19] was used to evaluate IRR. The index for agreement for the HCR-20 and non-HCR-20 studies was .97 and .98, respectively. The main source of disagreement was the study design, with discrepancies between raters arising when study authors described their design as 'retrospective,' but a careful perusal of the methods section revealed that the design in fact would be considered to be retrospective follow-up (or pseudo-prospective) in nature. The other types of disagreements most often resulted from a clerical error when entering item codes. Disagreements were discussed between raters, and the consensus coding was used in analyses in all but one case.[20]

[18] This is because the consensus ratings of the 2 raters (K. Reeves and N. Nikolova) against which IRR here was calculated were completed earlier for a different meta-analysis that focuses on HCR-20 research (Reeves et al., in preparation). Those consensus ratings were used for IRR purposes for the present project because the coding schemes were highly similar.

[19] Σ (agreements) / Σ (agreements + disagreements).

[20] Raters disagreed regarding the nature of the setting in which the violence occurred for one study (McGowan, 2007), and could not come to a consensus because the information presented in the study was highly unclear. Attempts made to obtain clarification from the study's author were unsuccessful. The rating of the IRR rater for this item was used in analyses.

Overlapping Studies and Datasets

When two or more disseminations examined the same dataset and no unique effect sizes could be calculated, the more methodologically sound dissemination was included (e.g., peer-reviewed publications were selected rather than unpublished sources on which they were based; studies that provided more descriptive information about relevant moderator variables were selected). Studies that clearly presented data on the same sample but from which unique effect sizes could be coded for any of the predictor, criterion, or moderator variables were included and coded as being from the same data set or sample. In several cases, the degree of sample overlap was unclear and could not be ascertained based on the published descriptions of the samples. In those instances, study authors were contacted to request clarification; additional information was obtained in all but two cases.[21]

Coding Rules and Procedures

Choice and Calculation of Effect Size

Although there are several types of effect sizes that can be estimated, each metric is conceptually synonymous and represents a standardized form of the magnitude of the observed effect (Field, 2003). Meta-analyses in the violence risk assessment literature have tended to use either correlations (e.g., Blair et al., in press; Bonta et al., 1998; Campbell et al., 2007; Edens & Campbell, 2007; Edens et al., 2007; Guy et al., 2005) or standardized mean differences (Cohen's *d*, Cohen, 1988; e.g., Hanson & Morton-Bourgon, 2005, 2007; Leistico, Salekin, DeCoster, & Rogers, 2008). Over the past several years, limitations inherent to correlations and standardized

[21] For these two sets of disseminations, the author's best judgment was used. Two studies first authored by C. Stadtland (Stadtland & Nedopil, 2005; Stadtland et al., 2005) were judged to present data on non-overlapping samples. Of four disseminations first authored by N. Gray, two were judged to present data from overlapping samples (Gray, Taylor, & Snowden, 2008; Gray, Fitzgerald, Taylor, MacCulloch, Snowden, 2007). Unique effect sizes were coded from each dissemination, and thus both were included in the meta-analysis and treated as a single sample. The two disseminations by Gray and her colleagues judged to present data on non-overlapping samples were Gray et al. (2003) and Gray et al. (2004).

mean differences within the context of investigating the predictive validity of violence risk tools have been described (Mossman, 1994; Rice & Harris, 1995). For example, interpretation of the magnitude of a correlation that represents the predictive validity of a risk assessment tool is affected by the base rate of violence - a variable that fluctuates widely in empirical studies in the violence risk literature - as well as the particular type of correlation coefficient being used. Cohen's *d* was intended for use under circumstances in which the scores being compared are both continuous and normally distributed, which, as noted by Rice and Harris (2005), are conditions often not met in the violence risk assessment field.

Receiver Operator Characteristic Analyses

Many researchers have recognized that certain characteristics of receiver operating characteristics (ROC) analyses appear to be particularly well suited for research in the violence risk assessment field (Mossman, 1994; Rice & Harris, 2005; Swets, Dawes, & Monahan, 2000). Mossman's (1994) meta-analysis, discussed earlier, is noteworthy for demonstrating that prediction accuracy is worthy of being studied in and of itself. Foreshadowing the frequent use of ROC analyses in subsequent research studies was Mossman's call to action that "future reports of prediction accuracy should use ROC methods and indices for quantifying results" (2004, p. p. 790). Although this analytic technique has been used in other fields such as radiology, radar signal detection, and sensory psychology for the past five or six decades (Swets, 1996), it did not emerge in research on violence risk assessment until approximately the past decade.[22] Whereas correlation coefficients are affected by variations in base rates and selection ratios, AUCs are not sensitive to these values. Moreover, ROC analyses were designed to be applied to data which comprise a continuous predictor variable and a dichotomous dependent measure, and therefore

[22] ROC analysis was developed within the context of signal detection theory during World War II for the analysis of radar images. Radar operators required support for their decision-making regarding whether to interpret a blip on the screen as an enemy target, an ally ship, or simply noise. Signal detection theory measures the ability of radar receiver operators to make such distinctions, and their ability to do so was termed "Receiver Operating Characteristics" (see Tape, n.d.).

are appropriate for research in the violence risk assessment field given the naturally occurring dichotomous state of the criterion of interest (i.e., violence occurs or does not occur during a given period of time). Given these characteristics, AUCs have been described as the "preferred measure of predictive or diagnostic accuracy in forensic psychology and psychiatry" (Rice & Harris, 2005, p. 618). Despite these certain benefits, ROC analyses are not without limitations, in that important temporal aspects relevant to the occurrence of violence, such as the frequency, imminence, and rate are not considered.

ROC analyses yield a plot of the true positive rate (sensitivity) against the false positive rate (1 minus specificity) for every possible cut-off score of the instrument. The term "receiver operating characteristic" (ROC) refers to its function of detecting (predicting) "characteristics" of the test, wherein the "receiver" of the data can "operate" at any point on the curve (Metz, 1978; Hanley & McNeil, 1982). Within the context of violence risk assessment research, the area under the curve (AUC) of the ROC graph can be regarded as an index for interpreting the overall accuracy of the predictor. Specifically, the AUC is the probability that a score on a risk assessment tool (assumed to be ordinal or continuous in nature) drawn at random from one sample or population (e.g., individual who perpetrated violence) will be higher than the score drawn at random from a second sample or population (e.g., individuals who did not perpetrate violence). AUC values can range from 0 (perfect negative prediction) to 1.0 (perfect positive prediction), with .50 indicating chance prediction. Values below .70 may be considered to be small; values in the range of .70 to .75 have been viewed as moderate; and values above .75 typically are regarded as large (see, e.g., Douglas, Yeomans, & Boer, 2005; Rice, 1997). However, such descriptors are tentative, as no definitive classification scheme exists (Bengtson & Långström, 2007; Sjösted & Långström, 2002), and, ultimately, establishing tolerable rates of error in prediction accuracy invokes consideration of societal and moral values (see, e.g., Bengtson & Långström, 2007; Buchanan & Leese, 2001; Monahan, 2001). In light of the above

characteristics of AUCs, they were chosen as the index of predictive accuracy for the present analyses. Since Mossman's (1994) pioneering meta-analysis in which he demonstrated meta-analytically the utility of using AUCs to assess different approaches to assessing risk for violence, only one other quantitative synthesis in the violence risk assessment field has used AUCs as the effect size metric of analysis (i.e., Schwalbe, 2007, who studied the predictive validity of risk assessment instruments used in juvenile justice settings).

For the present project, when a study reported an AUC value as well additional types of effect sizes for the same predictor/criterion pairing, (e.g., both AUC *and* correlation presented for the same type of outcome), only the AUC value was recorded. The majority (k[23] = 1205, 72.4%) of effect sizes included in the meta-analysis were AUCs reported by the studies' authors. Other types of effect sizes included were as follows: correlations (k = 402; 24.1%); d values calculated from means and standard deviations (k = 34; 2%); odds ratios calculated from frequency counts (k = 12; 0.7%); chi square values (k = 2; 0.1%); and standardized beta values (k = 10; 0.6%). When estimating an AUC value from a correlation, base rate information was considered when available (k = 249; 62% of correlations) using formulae based on Rosenthal (1991) and Swets (1986) that were summarized by Rice and Harris (2005). When transforming correlations to AUCs when base rate data were not available (k = 153; 38% of correlations), the base rate was assumed to be 50% (see Rice & Harris, 2005). Appendix 1.12 presents all transformational formulae used.

Weighting of Effect Sizes

Because an effect size estimate obtained from a large sample is assumed to be a more precise estimate of the population of effect sizes compared to an effect size based on a relatively small sample, effects typically are weighted such that those contributed by larger samples make a relatively more substantial "contribution." Although the practice of weighting effect sizes is

[23] k represents the number of effect sizes.

widely accepted in the meta-analytic literature, with the inverse variance typically considered to be a methodologically rigorous approach (Hedges & Olkin, 1985), researchers have recommended that meta-analyses be completed without weighted effect sizes (e.g., Rosenthal & DiMatteo, 2001) as well. In the present project, consistent with Schwalbe (2007), effect sizes were weighted by the study sample size. Such an approach has been deemed acceptable by experts in the meta-analytic field (see Lipsey & Wilson, 2001, pp. 36-37).

Dependence and Intercorrelation of Effect Sizes

All substantive analyses were conducted using independent effect sizes. That is, only one effect size from a sample (per level per predictor) was used in any given analysis. Several studies presented an effect size for an SPJ tool under two conditions: one based on the total sum of item scores, and another based on the structured, categorical summary risk rating. For analyses in which the overall validity of judgments made using an SPJ tool were examined, the average of these two effect sizes was used as the independent effect size for a given study.

Effect sizes presented by studies in which more than one predictor variable (risk assessment tool) was studied would be expected to be intercorrelated. Similar to the approach taken by other meta-analysts studying violence risk assessment (e.g., Hanson & Bussiére, 1998) or clinical and mechanical methods of prediction (Grove et al., 2000), these potential intercorrelations were ignored in the present project. The primary impact of this strategy is to yield a more conservative test of the differences between predictors (Hanson & Bussiére, 1998) that would not be expected to introduce a systemtic, predictable source of bias (Grove et al., 2000).

Missing Data

Reports of nonsignificance are problematic primarily because of the possibility that the results of the meta-analysis will be biased if zero or small effects are less likely than larger effects

to be reported in a study with the necessary information to compute the effect size. Although one strategy for handling this situation involves coding effects reported simply as "not significant" as zero, doing so is a conservative approach that downwardly biases the mean effect size across studies. When the primary research questions include "How big is the effect?" and "Is the effect different for different types of studies?"—questions that are of central interest in the present meta-analysis—Lipsey and Wilson (2001, p. 70) recommend "that analyses aimed at answering these latter questions not use effect sizes imputed as zero from reports of statistical nonsignificance." As such, the approach taken in the current project was to include only effects for which information necessary to calculate the actual value of the effect size was available. In an effort to examine the extent and seriousness of the "missingness" problem, the direction of effect for noncalculable effect sizes was examined.

There were 18 effect sizes, which represented less than one percent of all effect sizes coded, that were noncalculable. The 18 effect sizes were reported across 7 disseminations. Five of the 7 disseminations were retained in the meta-analysis because they presented effect sizes for other predictor/criterion pairings. These 5 studies accounted for 16 of the 18 noncalculable effect sizes. The 5 studies, along with values indicating the number of effect sizes included and excluded from analyses, respectively, were as follows: (a) Dernevik et al., (2002), 7 effect sizes coded, 4 excluded; (b) Meyers and Schmidt (2008), 10 included, 3 excluded; (c) Müller-Isberner, Sommer, Özokyay and Freese (1999), 15 included, 1 excluded; (d) Howard (2007), 3 included, 1 excluded; and (e) Lodewijks, de Ruiter, and Doreleijers (2008), 15 included, 7 excluded.

There were comparatively more reports of "not significant" than "significant." Of the 16 noncalculable effect sizes from the five disseminations that contributed effect sizes to analyses, there were three findings of "significant" and thirteen findings of "not significant." The direction of noncalculable effect sizes across predictor variables was as follows: PCL:SV: 1 not significant; Salient Factor Score: 1 not significant; HCR-20 Total: 1 not significant, 1 significant; HCR-20

Historical Scale: 1 not significant; HCR-20 Risk Management Scale: 1 significant; SAVRY Final Risk Rating: 3 not significant; SAVRY Total: 1 not significant; SAVRY Historical scale: 2 not significant; SAVRY Social/Contextual scale: 2 not significant; SAVRY Individual/Clinical scale: 1 not significant; and SAVRY Protective Factors scale: 1 not significant.

In only two cases did this approach to handling missing data regarding statistical significance of effect sizes result in the exclusion of a study. Both studies excluded from the meta-analysis on this basis (Levene et al., 2004; Yuille, 2008) investigated the EARL-21G. The findings regarding predictive validity from the two studies were in opposite directions (i.e., Levene et al., 2004, reported "significant" findings; Yuille (2008) reported findings as "not significant"). As such, there is no evidence that the pattern of missingness was biased toward one particular direction of significance, at least for studies that presented noncalculable effect sizes and therefore were not included in the meta-analysis.

Overall, there does appear to have been a bias in terms of greater reports of nonsignificance. Given that the 18 noncalculable effect sizes represented less than one percent of all effect sizes coded, it is highly unlikely that the missing data had any impact on the meta-analytic findings.

Choice of Statistical Model

Two main statistical models have been developed to make inferences about the average effect size in meta-analysis: fixed-effects and random-effects models (a hybrid approach, the mixed-effects model, also exists; see Hedges, 1992). The major difference between the two models is that fixed-effects models assume that sample effect sizes are homogeneous and estimate a single population parameter, whereas random-effects models assume that sample effect sizes are heterogeneous and sample from a distribution of population effect sizes (see Hedges, 1992; Hedges & Vevea, 1998; Hunter & Schmidt, 2000).

Choice of statistical model should be guided by the inferences that the meta-analyst wishes to make (Hedges & Vevea, 1998). Fixed-effects models are appropriate for making conditional inferences, which can be made only about the effect size parameters in the observed set of studies. More specifically:

> Strictly speaking, conditional inferences apply to *this* collection of studies and say nothing about other studies that may be done later, could have been done earlier, or may have already been done but are not included among the observed studies. (Hedges & Vevea, 1998, p. 487)

On the other hand, a random-effects model facilitates drawing inferences about the parameters of the population of studies that go beyond the set of observed studies and that would not be expected to be identical to them (i.e., unconditional inferences). The observed studies are of interest only to the extent that "they reveal something about a putative population of studies that are the real object of inference" (Hedges & Vevea, 1998, p. 487).

Another way in which the models differ is in the calculation of the weights used in the analysis. This difference influences the standard errors associated with the aggregate effect size. Because fixed-effects models assume all "unknowns" in the model to be constant, they only use within-study variability in their weights (Hedges & Vevea, 1998). In contrast, random-effects models account for the errors associated with sampling, and population effect sizes can be thought of as being sampled from a universe of possible effects. The error term in a random-effects model contains both between and within study variability. The consequence of having a larger standard error term renders the significance test of combined effects more conservative in the random-effects model compared to the fixed-effects model (Field, 2003).

The differences in the models have prompted many experts in meta-analytic methodology to recommend the use of random-effects analyses instead of fixed-effects analyses (e.g., Field, 2003; Hunter & Schmidt, 2000). Of course, random-effects analyses are not without limitations. Cohn and Becker (2003) commented that most meta-analyses that use random-effects analyses

violate a fundamental assumption of the random-effects model because the studies included in a meta-analysis probably never represent a random selection of studies that have been or could have been conducted in a research domain. Nevertheless, given the distinct advantages that random-effects models confer over fixed-effects model, coupled with the intent of the present project to permit unconditional inferences to be drawn, the present meta-analysis will be completed using a random-effects model. More specifically, moderator analyses (see below) were conducted using a maximum-likelihood random effects model.

Choice of Statistical Tests and Analyses

Weighted effect sizes were aggregated using a meta-analysis macro for SPSS written by David B. Wilson (see Lipsey & Wilson, 2001; available online at http://mason.gmu.edu/~dwilsonb/ma.html) that is intended for use with any type of effect size. The macro calculates upper and lower 95% confidence intervals (CIs) and performs a homogeneity test using the Q statistic (Hedges & Olkin, 1985), which is distributed as a chi-square with $k - 1$ degrees of freedom, where k is the number of studies (or independent effect sizes contributed by all studies). In the event that considerable heterogeneity is identified, moderator variables should be examined statistically to identify possible sources of this variability.

Differences between groups of effect sizes were examined using Lipsey and Wilson's (2001) "MeanES" macro. Analyses of categorical moderator variables were conducted using their "MetaF" macro, which is a meta-analytic analog to an analysis of variance. This macro divides the total variability among observed effect sizes into the portion that can be explained by the categorical predictor variable (Q_B) and the residual portion (Q_W). A statistically significant Q_B value indicates that the difference between the means of the categories is greater than what would be expected based simply on sampling error (Lipsey & Wilson, 2001, pp. 135–136).

RESULTS

Study Characteristics

As of August 9, 2008, 115 disseminations that comprised 106 independent samples had been identified, retrieved, and coded, yielding 1881 effect sizes. Subsequent to applying coding rules related to findings reported as 'not significant' as detailed above, 1665 effect sizes from 113 disseminations that represented 104 non-overlapping samples remained for analyses. Each sample contributed on average 16 effect sizes (*SD* = 18.93; range: 1–120). These studies represented a total nonredundant sample size of 14,638 cases (*M*=139, *SD*=167, range= 6-1465).

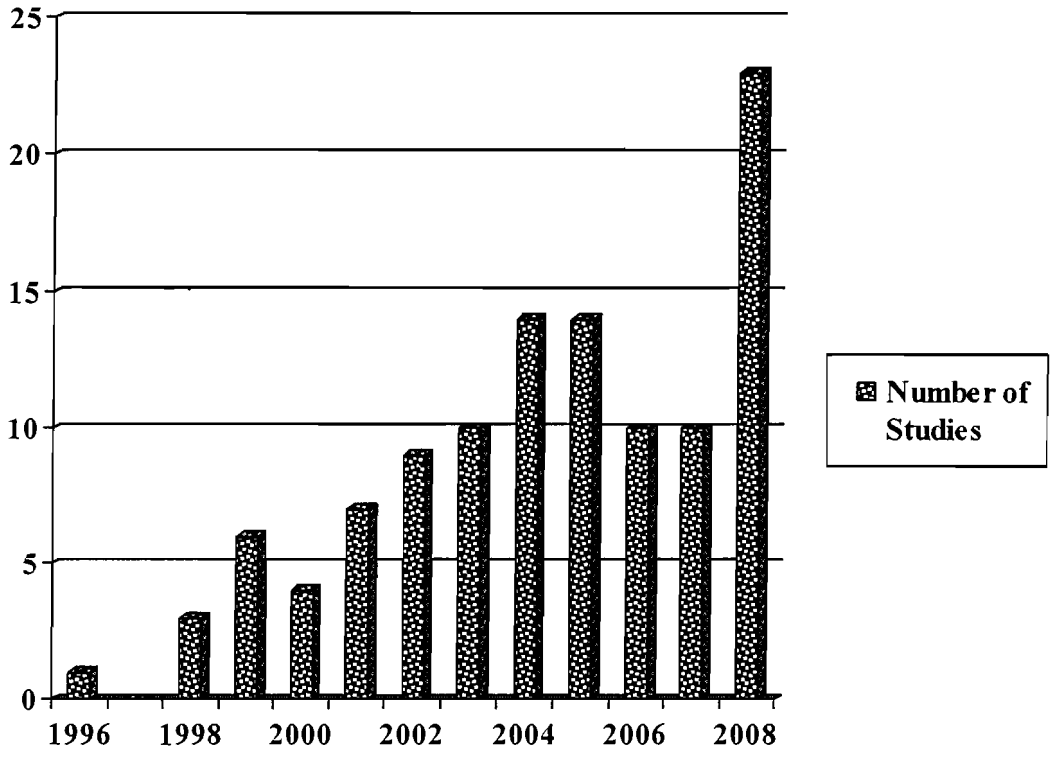
Studies were completed between 1996 and 2008, with the median year being 2005 (see Figure 1.1). Most disseminations (69; 61.1%) were retrieved from published peer-reviewed journals. Other data sources included doctoral dissertations (15; 13.3%), master's theses (3; 2.7%), conference presentations (18; 15.9%); manuscripts or technical reports retrieved from non-peer-reviewed sources (7; 6.2%); and one book chapter (.9%). The country where most studies were completed was Canada (36; 31.9%), although the majority of studies was carried out in Europe (59; 52.2%), with 18 (15.9%) studies having been completed in the United States.

Of the 113 disseminations, four were coded with the assistance of translators,[24] and four were coded based on English summaries of the research project.[25]

[24] M. Collins translated two studies from French (Pham, Chevrier, Nioche, Ducro, & Réveillère, 2005; Pham, Ducro, Marghem, & Réveillère, 2005), and L. Cuadra translated two studies from Spanish (Folino, Almirón, & Ricci, 2007; Ramírez, Illescas, García, Forero, & Pueyo (2008). Both individuals who acted as translators have completed research in the violence risk assessment field, and are fluent in both English and the non-English language.

[25] Hildebrand et al. (2005); Neves and Gonçalves (2008); Schönberger et al. (2008); and Stadtland and Nedopil (2005).

Figure 1.1. Number of Studies Completed by Year



Notes. Total number of studies is 111. Two studies are not represented in the histogram because they are unpublished manuscripts whose year of completion was unknown.

Among the disseminations that reported mean age (98 of 113), the average of the mean ages was 31.14 years (SD = 9.49; range = 9.60 – 50.15). Most studies (91; 80.5%) reported on adults, with the remainder (20 studies; 17.7%) reporting on juveniles and children (2 studies; 1.8%). Although most samples consisted of men (59; 52.2%), a large number (48; 32.4%) reported on mixed gender samples. There were six samples (5.3%) that comprised only women. Seven[26] of the 104 samples (6.7%) contributed effect sizes for both women and men separately. An alarming majority of studies (54%) did not present information regarding the racial/ethnic composition of the sample. These missing data possibly may exist because many studies

[26] Cooper, Eslea, & Ireland (2008); de Vogel & de Ruiter (2005); Fitch (2002); Lodewijks, de Ruiter, & Doreleijers (2008); Meyers & Schmidt (2008); Nicholls (2001); and Nicholls et al. (2004).

completed in Europe and Canada likely would have had relatively homogeneous samples with little variability, and thus potentially would have been less likely to have been reported. Among samples for whom racial/ethnic data were available, 67% were described as Caucasian or White.

Mean base rates across effect sizes from all studies varied as a function of outcome category: Any antisocial behaviour (k = 259; M = 40.89%; SD = 18.11; range: 5.23 – 89.70); Violence (k = 149; M = 30.37%; SD = 20.16; range: 5.00 – 30.37); Physical violence, excluding sexual violence (k = 109; M = 28.16%; SD = 10.72; range: 12.70 – 56.00); Physical violence, including sexual violence (k = 412; M = 33.54%; SD = 7.66; range: 5.3 – 93.00); Sexual violence (k = 109; M = 19.83%; SD = 10.93; range: 8.30 – 57.41); Intimate partner violence (k = 35; M = 26.04%; SD = 11.58; range: 7.97 – 59.80); and Nonviolent antisocial behaviour (k = 213; M = 43.30%; SD = 21.27; range: 10.00 – 92.00). Of the 104 samples, the mean time at risk was not specified for 18 samples. The average length of time that participants were at risk in the remaining 86 samples was 37.38 months.[27] For most samples, outcome data were recorded from a single source. In 18 eighteen studies, outcome data were gathered from 2 sources, in 7 studies date were gathered from 3 sources, and in 1 study data were gathered from 4 sources. Most sources of data were recorded from official criminal justice records (62.39%). Hospital records (29.06%), staff observation (3.42%), self-report (2.56%), and collateral sources, usually family members, (2.56%) represented the other sources of information.

In total, of all 1665 effect sizes, 956 represented an index from a SPJ tool (total scores, k = 282; summary risk ratings, k = 83; scale scores, k = 584; actuarially derived categorical risk categories, k = 7). In terms of actuarial indices, 253 and 81 effect sizes were coded for total and scale scores, respectively, for actuarial measures. There were 153 effects coded for the total score of any of the PCL family of measures, with an additional 204 that corresponded to PCL scale

[27] Among the 59 samples in which the standard deviation of the mean length of follow-up was reported, the average standard deviation was 11.23 months. Among the 35 studies in which ranges were provided, the average minimum length of follow-up was 16.42 months, and the corresponding average maximum length was 94.69 months.

scores. Finally, 18 effect sizes were presented for judgments made using unstructured clinical judgment. These values represent non-independent effect sizes, wherein as many effect sizes as possible were coded from each study (e.g., the $k = 83$ effect sizes for summary risk ratings represent effects for the same risk index across multiple types of outcomes within a given study, such as, for example, effects coded from a single sample for SAVRY summary risk ratings corresponding to sexual violence and additional, separate effects corresponding to non-sexual physical violence.)

At the *dependent* one-risk-index-per-sample level, there were 32 effect sizes coded for summary risk ratings using an SPJ tool,[28] and 110 effects for SPJ total scores.[29] In total then, for either type of SPJ index (i.e., numeric or summary risk rating), across all 104 samples, 142 effect sizes were coded. For studies that presented effect sizes for both the summary risk rating and the total score of the same measure, the average of these two values was used to represent the overall SPJ index at the *independent* one-risk-index-per-sample level.

Disseminations included in the meta-analysis are marked with an asterisk in the references. When effect sizes were collapsed (averaged) at the individual study level, one effect for an SPJ tool was observed for each study (per inclusion criteria), and, additionally, there were 45 actuarial effect sizes, 55 PCL effect sizes, and 6 unstructured clinical judgment effects.

[28] Each sample that contributed an effect size for SPJ summary risk rating did so only for one SPJ tool except for two studies (Hildebrand et al., 2005; Schönberger et al., 2008) that reported on two SPJ measures.

[29] Each sample contributed one effect size for SPJ total score except for the following 7 studies that reported on two SPJ measures: Allen and Howells (2008); Grann and Wedin (2002); Grann et al. (2005); Hildebrand et al. (2005); Hill, Habermann, Klusmann, Berner, & Briken (2008); Schönberger et al. (2008); Stadtland et al. (2005).

Performance of the SPJ Approach

Central Tendency and Distribution

Table 1.1 summarizes the central tendency and distribution of raw (not weighted) and weighted AUC values for assessment of risk using an SPJ tool across: (a) all 949 effect sizes, (b) one effect size per SPJ index (yotal score or summary risk rating), and (c) one effect size per dataset. Findings presented at the level of one effect size per dataset represent the average within study judgments of risk based on total scores and summary risk ratings from all SPJ tools used in a particular study. A stem-and-leaf plot and histogram are presented in Figures 1.1 and 1.2, respectively, to assist in visualizing the shape of the distribution of the 949 effect sizes.

As Table 1.1 and Figures 1.2 and 1.3 indicate, the distribution of effect sizes closely approximated a normal curve. There was some dispersion of AUCs around the median when all 949 effect sizes that represented all SPJ tool indices (i.e., total scores, summary risk ratings, and scale scores) were considered. Considering only the statistically independent effect sizes (i.e., one AUC per sample), a larger degree of dispersion was observed, with the middle fifty percent of effect sizes (AUCs) from the 104 samples falling between .63 and .76. Thus, at least 25% of all the independent effect sizes were moderate to small in magnitude, whereas about 25% of the AUCs were quite large in magnitude.

Table 1.1. Association between SPJ Risk Assessment and Antisocial Behaviour: Distribution of AUC Values

Assessment Approach	*k*	*n*	Not Weighted				Weighted		
			AUC (SE)	*SD*	Mdn	Q1–Q3	AUC_w (SE)	95%CI_w	*Q*
All AUCs	949	–	.66 (.01)	.11	.66	.59-.73	.65 (.003)	.65-.66	1198.86***
Summary risk rating	83	–	.68 (.01)	.11	.68	.59-.75	.68 (.01)	.65-.70	105.79*
Total scores	282	–	.68 (.01)	.11	.68	.62-.76	.68 (.01)	.66-.69	336.10**
Scale scores	584	–	.65 (.01)	.11	.64	.57-.71	.64 (.004)	.63-.65	718.43***
Any summative AUC per Risk Index per Sample	143	–	.69 (.01)	.10	.70	.63-.76	.68 (.01)	.67-.70	142.63
Summary risk rating	32	3,595	.69 (.02)	.10	.71	.63-.74	.69 (.02)	.65-.72	27.01
Total scores	110	15,514	.70 (.01)	.10	.69	.62-.76	.68 (.01)	.67-.70	115.58
One AUC per Sample^	104	15,077	.70 (.01)	.10	.70	.63-.76	.68 (.01)	.67-.70	114.17

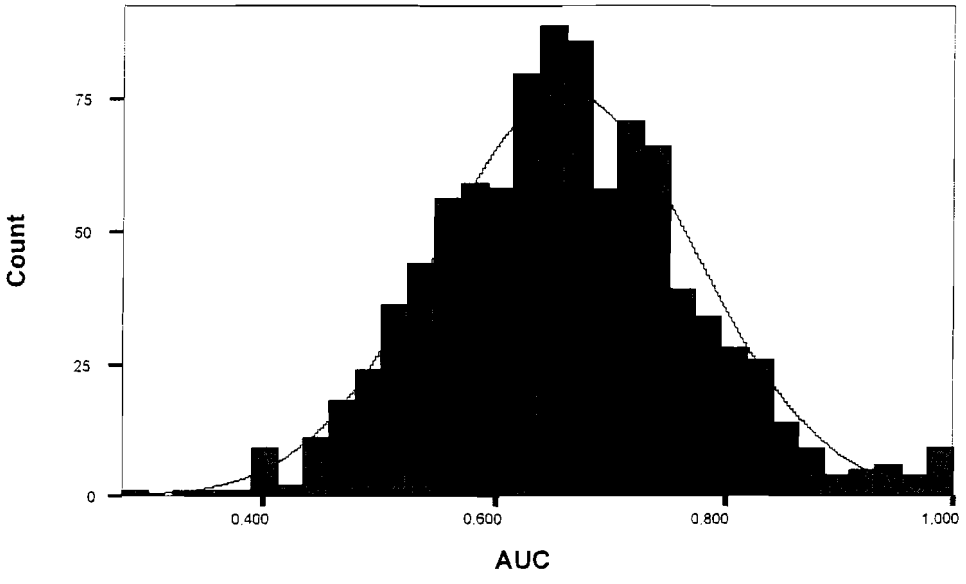
Note. Criterion is any antisocial behaviour. *k* = number of effect sizes; SE = standard error of the mean; Mdn = Median; Q1 = first quartile; Q3 = third quartile. AUC_w(SE)= mean weighted effect size (AUC) and standard error of the weighted mean effect size; 95% CI_w = 95% confidence interval of mean weighted effect size. Summary risk rating refers to structured, categorical ratings of risk (i.e., low, moderate, or high). Effect sizes (AUCs) in the first and second horizontal layers are not independent. ^ Represents the average within sample judgments of risk based on total scores and summary risk ratings from all SPJ tools used in a particular sample. *$p \geq .05$. **$p \geq .01$. ***$p \geq .001$. Seven of the 1665 effect sizes represent numerical (i.e., actuarially based) SPJ risk categories and were excluded from these analyses. The sample size presented for $k = 110$ total scores (15,514) also contains individuals counted in the $k = 32$ summary risk ratings (3,595) because some studies presented effect sizes for both SPJ numeric and summary risk rating indices.

Figure 1.2. Stem-and-Leaf Plot of the 949 SPJ-based AUC Values

Frequency	Stem	& Leaf
3.00	Extremes	(=<.36)
5.00	3.	9&
11.00	4.	04&
40.00	4.	5567788999
90.00	5.	0001111222223333444444
122.00	5.	555555666667777888888899999999
163.00	6.	000000111111222222222333333333344444444
175.00	6.	55555555566666666777777777788888888888899999
153.00	7.	00000001111111122222222233333334444444
83.00	7.	55566667777788889999
53.00	8.	000011223344
26.00	8.	5566789
9.00	9.	3&
16.00	Extremes	(>=.95)

Note. & = fractional leaves. Stem width = .100. Each leaf = 4 cases.

Figure 1.3. Histogram of the 949 SPJ-based AUC Values



The heterogeneity of the effect sizes was evaluated by calculating Q, following Shadish and Haddock (1994; p. 266). For all AUCs, the obtained value of Q was 1198.86 ($df = 948$, $p <$ .0001), indicating substantial variability among effect sizes. Q remained statistically significant

when only effect sizes for summary risk ratings were considered ($Q = 105.79$, $df = 82$, $p = .04$), when only total scores were considered ($Q = 336.10$, $df = 281$, $p = .01$), and when only scale scores were considered ($Q = 718.43$, $df = 583$, $p < .0001$).

When effect sizes at the level of any summative risk index per sample (i.e., SPJ total score and summary risk rating, with both types of effects contributed by the same study when available) were evaluated, Q was not significant ($Q = 142.63$, $df = 142$, $p = .47$). This remained the case irrespective of whether effect sizes for only summary risk ratings ($Q = 27.01$, $df = 32$, $p = .72$) or total scores ($Q = 115.58$, $df = 109$, $p = .32$) were considered. Similarly, when only statistically independent effect sizes were evaluated (i.e., one SPJ index per sample) Q remained not significant ($Q = 114.17$, $df = 103$, $p = .21$).

Although these findings suggest the distribution of effect sizes to be homogeneous, the observed variability at the one-per-study level does in fact exceed expectation, but merely not to such a degree as to be statistically significant at $p < .05$. More specifically, because the expected value of a chi-square is equal to its degrees of freedom, and the observed one-per-study Q value of 114.17 is larger than its 103 associated degrees of freedom, the observed variability exceeds what would be expected by chance (see Lipsey & Wilson, p. 135). These analyses indicate that, notwithstanding the non-significance of Q for some tests, there was considerable variability with respect to the magnitude of effect sizes. These findings, in conjunction with theoretical and empirical rationale for analyzing *a priori* identified variables as outlined in the Introduction, support the search for factors that potentially moderate the association between antisocial behaviour and risk assessments made using the SPJ approach (see below).

Publication Bias

Publication bias is an unavoidable issue to be addressed in meta-analysis. This bias refers to the assumption that studies in which positive findings are reported are more likely to be

published than studies in which null results are reported. The implication of the existence of such a bias would be a systematic over-estimation of the true magnitude of effect sizes reported in the published scientific literature. Although great efforts to minimize any potential publication bias were made through a comprehensive search strategy that included efforts to retrieve unpublished research in the form of dissertations, theses, conference presentations, unpublished manuscripts, and government or business reports not subjected to peer-review, it is inevitable that at least some relevant unpublished studies were not included. However, nearly half (k = 42; 40.38%) of all samples included in the present project were retrieved from sources other than peer-reviewed journals.

Analyses to investigate whether the difference between the weighted mean effect size of studies retrieved from peer-reviewed published journals versus the weighted mean effect size of studies retrieved from other sources such as unpublished dissertations, theses, manuscripts, government reports, and conference presentations was greater than what would be expected simply by sampling error were not significant: Q_B (df = 1) = .08, p = .78 (see Table 1.2). The mean weighted AUC values of these two groups was virtually indistinguishable (AUC_w = .69 and .68, respectively) when the outcome of any antisocial behaviour was studied. The pattern of nonsignificant findings was observed across all seven categories of the criterion.

Table 1.2. Publication Status

Publication Type	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Journal Article	62	10,202	.69 (.01)	.66-.71	42.48
Other^	42	4771	.68 (.02)	.65-.71	46.18

Note. k = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. For overall test, n = 14,973; k = 104; AUC_w = .68 (SE = .01); 95% CI = .67-.70; Q_B (df = 1) = .08, p = .78; Q_W (df = 102) = 88.66, p = .82; Z = 71.44, $p \leq .001$. All levels of this moderator have weighted mean effect sizes that differed significantly from chance, as indicated by a z-test ($p \leq .001$). ^ Comprises conference presentations, dissertations, theses, unpublished manuscripts, and government reports not subjected to peer-review.

Application of SPJ Tools across Different Forms of Antisocial Behaviour

The extent to which the type of risk judgment made using an SPJ tool—that is, summing the items to yield a numeric score versus rendering a final structured, categorical[30] summary risk rating—may differ in magnitude of predictive validity as a function of the type of outcome observed was examined. Mean weighted effect sizes for both types of judgments across the seven categories of antisocial behaviour were computed using the Lipsey and Wilson's (2001) 'MeanES' SPSS macro. In addition, because a given study could contribute effect sizes for both types of risk judgments for the same measure, thereby creating statistical dependency among the effect sizes, Table 1.3 also presents statistically independent findings for SPJ judgments in which only one effect size was contributed per sample (i.e., the average of the AUCs for numeric score and summary risk rating, when both indices were presented in a given sample).

As can be seen in Table 1.3., for effect sizes at the one per sample level, mean weighted AUCs ranged from .59 for sexual violence to .74 for violence, with the mean weighted effect size for the other five outcome categories being either .65 or .68. Across the seven types of violence, the mean weighted AUC value for the summary risk rating was larger in magnitude than the mean weighted effect for the numeric score for all but two outcome categories. Typically, the differences between the numeric indices and summary risk ratings were relatively small.

None of the Q values was statistically significant (although a strong trend towards significance was observed for SPJ total scores in the nonviolent aggression category, $p = .06$), which indicates that there was little to no variability across types of outcome categories. Therefore, with the exception of certain exploratory analyses, the moderator analyses reported below are presented only for the outcome category of any antisocial behaviour.

[30] Of course, a structured professional judgment regarding an individual's risk for violence need not necessarily be communicated using a 3-level category; it is studied as such herein though because all known research has used this procedure.

Table 1.3. SPJ Approach and Type of Antisocial Behaviour

SPJ Index	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Any Antisocial Behaviour					
Aggregate^	104	14,973	.68 (.01)	.67 - .70	109.84
Total score	93	14,155	.68 (.01)	.66 - .70	104.88
Summary Risk Rating	32	3583	.69 (.02)	.65 - .72	26.75
Violent					
Aggregate	25	2530	.74 (.02)	.70 - .77	20.23
Total score	22	2238	.74 (.02)	.70 - .78	18.92
Summary Risk Rating	10	969	.71 (.03)	.64 - .77	7.02
Physical (Including Sexual)					
Aggregate	47	6356	.68 (.01)	.66 - .71	47.95
Total score	40	5659	.68 (.01)	.65 - .70	39.14
Summary Risk Rating	9	985	.77 (.04)	.71 - .84	8.81
Physical (Excluding Sexual)					
Aggregate	11	220	.65 (.03)	.59 - .70	6.33
Total score	9	1044	.66 (.03)	.60 - .72	5.34
Summary Risk Rating	9	988	.63 (.03)	.57 - .69	6.31
Sexual					
Aggregate	14	1682	.59 (.04)	.52 - .66	26.92
Total score	14	1590	.60 (.04)	.52 - .67	26.21
Summary Risk Rating	9	827	.65 (.04)	.57 - .72	11.54
Intimate Partner Aggression					
Aggregate	8	3264	.65 (.02)	.62 - .68	4.91
Total score	8	3602	.63 (.02)	.60 - .67	5.99
Summary Risk Rating	4	563	.73 (.04)	.64 - .81	1.79
Non-violent					
Aggregate	35	3752	.65 (.02)	.62 - .69	40.63
Total score	27	2940	.65 (.02)	.60 - .70	38.09
Summary Risk Rating	12	1313	.66 (.03)	.60 - .71	9.15

Note. *k* = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. Levels of all variables have weighted mean effect sizes that differed significantly from chance, as indicated by a *z*-test ($p \leq .001$). ^ Data in the first row of each layer ('Aggregate') represent independent effect sizes (i.e., one effect size per sample), which is the aggregate of the within study average of risk assessments based on SPJ total and/or Summary Risk Rating.

Moderator Analyses

Variables that directly impact clinical practice and those that inform the way in which empirical research is (or should be) carried out were examined for any potential impact on the predictive validity of estimates of risk made using an SPJ measure. All moderators variables

examined were categorical in nature. Therefore, moderator analyses were carried out using Lipsey and Wilson's (2001) 'MetaF' macro, a meta-analytic analogue to an ANOVA, to examine the difference among levels of moderators.

Factors Affecting Clinical Practice

Gender

Most studies presented findings for mixed gender samples or samples comprising only men. For studies that reported on mixed gender samples, it typically was the case that separate effects could not be estimated for women and men. Seven samples contributed effect sizes for both women and men.[31]

Studies that used only men, only women, or both men and women as participants did not produce effect sizes that differed from one another at a statistically significant level. However, the value for Q_B was small (Q_B (df = 2) = 5.57) and approached significance (p = .06), with the 95% confidence intervals for women and men overlapping by only one decimal point (i.e., AUC_w = .70). Specifically, there was a trend for effect sizes to be greater among samples comprising only women (AUC_w = .78) than only men (AUC_w = .68). Effect sizes were significantly different than chance at each level of this moderator (see Table 1.4.).

Table 1.4. Gender

Gender	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Journal Article	60	9809	.68 (.01)	.66 - .70	48.09
Other^	10	653	.78 (.04)	.70 - .86	12.41

Note. k = number of effect sizes; AUC_w = mean weighted effect size (AUC) ; SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. For overall test, k = 111; AUC_w = .69 (SE = .01); 95% CI = .67 - .71; Q_B (df = 2) = 5.57, p = .06; Q_W (df = 108) = 94.64, p = .82; Z = 74.22, $p \leq .001$. All levels of this moderator have weighted mean effect sizes that differed significantly from chance, as indicated by a z-test ($p \leq .001$).

[31] Cooper et al. (2008); de Vogel & de Ruiter (2005); Nicholls et al., (2004); Fitch (2002); Lodewijks, de Ruiter, et al. (in press); Meyers & Schmidt (2008); Nicholls (2001).

Age

Irrespective of whether participants were adults (18 years of age and older), adolescents (between 13 and 17 years of age), or children (12 years of age and younger), mean weighted effect sizes were not statistically significantly different: Q_B (df = 2) = 3.20, p = .21 (Table 1.5). However, the mean effect size for children (AUC_w = .59) was based on only two AUC values. Moreover, the weighted effect size for children was appreciably smaller in magnitude than the values for adults (AUC_w = .69) and adolescents (AUC_w = .68), and generated a 95% confidence interval that did not overlap with that for adults.

The pattern of nonsignifcant findings was observed across all outcome categories except for physical (including sexual) violence: Q_B (df = 1) = 3.70, p = .05. That analysis included only adults and adolescents and indicated that weighted effect sizes were larger among adolescents (AUC_w = .75 ; 95% CI : .68 - .82) than adults (AUC_w = .67 ; 95% CI : .65 - .70) when physical (including sexual) violence served as the criterion.

Table 1.5. Age Group

Age Group	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Adult	83	12,612	.69 (.01)	.67 - .71	74.32
Adolescent	19	1906	.68 (.02)	.63 - .73	15.40
Child	2	455	.59 (.05)	.48 - .70	1.18

Note. k = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. For overall test, n = 14,973; k = 104; AUC_w = .68 (SE = .01); 95% CI = .67 - .70; Q_B (df = 2) = 3.20, p = .21; Q_W (df = 101) = 90.90, p = .75; Z = 74.14, $p \leq .001$. All levels of this moderator have weighted mean effect sizes that differed significantly from chance, as indicated by a z-test ($p \leq .001$).

Nationality

To examine whether nationality moderated predictive accuracy of SPJ judgments of violence risk, studies were coded according to whether data were collected in Europe or North America. The meta-analytic analogue to the ANOVA was not significant: Q_B (df = 1) = 2.80, p = .09 (Table 1.6).

In light of previous meta-analytic findings (Guy et al., 2005) that predictive validity of the PCL-R and PCL:SV were moderated by country of data collection, with studies conducted in the United States obtaining much weaker effects than studies conducted in Canadian and European settings, a similar tri-categorization of nationality was evaluated in the current project as well. Similar to results for the first set of analyses that compared European and North American studies, no significant differences were observed when North American studies were split into studies completed in Canada or the United States: Q_B (df = 2) = 2.80, p = .25.

Table 1.6. Country of Data Collection

Continent	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Europe	56	6981	.70 (.01)	.67 - .73	35.31
North America	48	7992	.67 (.01)	.64 - .69	53.50

Note. k = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. For overall test, n = 14, 973; k = 104; AUC_w = .68 (SE = .01); 95% CI = .67 - .70; Q_B (df = 1) = 2.80, p = .09; Q_W (df = 102) = 88.81, p = .82; Z = 72.87, $p \leq .001$. Both levels of this moderator have weighted mean effect sizes that differed significantly from chance, as indicated by a z-test ($p \leq .001$).

Clinical Setting/Population

Four types of population were examined: civil psychiatric patients, forensic psychiatric patients, correctional offenders (i.e., parolees, probationers, inmates), and mixed groups. The mixed categories comprised samples that consisted of civil and forensic psychiatric patients (e.g., Doyle & Dolan, 2006), correctional offenders and forensic psychiatric patients (e.g., Penney, Lee, Moretti, & Bartel, 2007), persons in the general population, including victims (e.g., Heckert & Gondolf, 2004), and an educational facility for emotionally disturbed students (McGowan, 2007).

As indicated in Table 1.7., most studies reported on samples of forensic psychiatric (k = 51) or correctional offenders (k = 35), with civil psychiatric (k = 5) and mixed (k = 13) samples accounting for the remaining categories. Across the four types of populations, mean weighted effect sizes did not differ: Q_B (df = 3) = 1.87, p = .60.

Setting of Violence

Most research investigated the predictive validity of SPJ tools for assessing risk for violence in the community ($k = 76$), with 29 effects analyzed for institutions and 3 for mixed settings. Four studies[32] contributed an effect size for two levels (institutional and community) of this moderator analysis.

Table 1.7. Clinical Setting/Population

Setting	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Civil	5	580	.71 (.05)	.62 - .79	1.20
Forensic	51	6694	.67 (.01)	.64 - .70	35.29
Correctional	35	6010	.70 (.02)	.67 - .73	48.27
Mixed/Other	13	1689	.69 (.03)	.64 - .75	3.62

Note. k = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. For overall test, n = 14,973; k = 104; AUC_w = .69 (SE = .01); 95% CI = .67 - .70; Q_B (df = 3) = 1.87, p = .60 ; Q_W (df = 100) = 88.67, p = .79; Z = 72.19, $p \leq .001$. All levels of this moderator have weighted mean effect sizes that differed significantly from chance, as indicated by a z-test ($p \leq .001$).

As Table 1.8 indicates, the strength of the association between SPJ risk assessment and antisocial behaviour did not differ as a function of whether violence was measured within an institution (i.e., psychiatric hospital or correctional facility; AUC_w = .67) or in the community (AUC_w = .69; Q_B (df = 2) = .99, p = .62). Effect sizes were significantly greater than chance at all levels of this moderator.

Information Used to Assess Risk

Effect sizes were examined as to whether their magnitude differed as a function of the information sources used to complete the risk assessment. In most studies, the SPJ tool was completed solely on the basis of file information ($k = 55$), with 33 samples reporting findings for risk assessments based on both file and interviews. Studies that used a pseudo-prospective design

[32] Cooke, Michie, & Ryan (2001); Dernevik (2004); Dernevik et al. (2002); and Viljoen et al. (2008).

and reported having recorded scores of a risk measure from files that originally had been completed using files and an interview were assigned to the 'file plus interview' level of the moderator. The mixed/other category represents, for example, cases in which only an interview was used, or cases in which one part of the measure were completed using files only and another part, interview only (e.g., scoring the HCR-20 Historical scale from files, but the other scales based on interviews).

Table 1.8. Location of Antisocial Behaviour

Location	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Institution	29	2678	.67 (.02)	.63 - .71	15.41
Community	76	12,503	.69 (.01)	.67 - .71	75.92
Mixed	3	203	.73 (.07)	.59 - .88	.35

Note. *k* = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. For overall test, *n* = 15, 384; *k* = 108; AUC_w = .68 (SE = .01); 95% CI = .67 - .70; Q_B (*df* = 2) = .99, *p* = .62 ; Q_W (*df* = 105) = 91.68, *p* = .82; *Z* = 72.84, $p \leq .001$. All levels of this moderator have weighted mean effect sizes that differed significantly from chance, as indicated by a *z*-test ($p \leq .001$).

As Table 1.9 indicates, whether a risk assessment using an SPJ tool was completed on the basis of files alone (AUC_w = .68) or supplemented by an interview (AUC_w = .70), the mean weighted effect size did not vary significantly across groups: Q_B (*df* = 2) = .67, *p* = .72.

Table 1.9. Source of Information Used to Complete Risk Assessment

Source	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Files only	55	8247	.68 (.01)	.65 - .70	47.04
Files + Interview	33	3389	.70 (.02)	.66 - .73	34.25
Mixed/Other	16	3337	.69 (.02)	.65 - .73	7.68

Note. *k* = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. For overall test, *n* = 14, 973; *k* = 104; AUC_w = .68 (SE = .01); 95% CI = .67 - .70; Q_B (*df* = 2) = .67, *p* = .72 ; Q_W (*df* = 101) = 88.97, *p* = .80; *Z* = 71.89, $p \leq .001$. All levels of this moderator have weighted mean effect sizes that differed significantly from chance, as indicated by a *z*-test ($p \leq .001$).

Factors Affecting Research

Design

Study design was a statistically significant moderating factor of the magnitude of effect size: Q_B ($df = 2$) = 6.29, p = .04 (see Table 1.10). Studies that reported on the post-dictive validity of an SPJ measure generated larger weighted effect sizes (AUC_w = .75) compared to studies that reported on predictive validity using either a true prospective (AUC_w = .67) or pseudo-prospective (AUC_w = .68) design.

Table 1.10. Research Design

Design	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
True Prospective	34	3969	.67 (.02)	.64 - .71	28.68
Pseudo-prospective§	54	9674	.68 (.01)	.65 - .70	41.79
Retrospective	16	1330	.75 (.03)	.70 - .81	17.93

Note. k = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. For overall test, n = 14,973; k = 104; AUC_w = .68 (SE = .01); 95% CI = .67 - .70; Q_B ($df = 2$) = 6.29, p = .04 ; Q_W ($df = 101$) = 88.41, p = .81; Z = 74.45, $p \leq .001$. All levels of this moderator have weighted mean effect sizes that differed significantly from chance, as indicated by a z-test ($p \leq .001$). § Also referred to a retrospective follow-up.

Allegiance

Analyses undertaken to examine the degree to which a study's author(s) was affiliated with the risk measure under investigation in that study were completed using effect sizes that were not statistically independent. Because almost half of the samples (k = 49; 47%) presented an effect size for more than one risk assessment measure, effects were retained for all measures presented in a given study despite statistical dependence. This approach is consistent with the method followed by Marcus et al. (in press) in their meta-analytic investigation on this topic. For the 27 samples in the present study that reported an effect size for both the total score and the summary risk rating of the same SPJ tool, the average of the two effect sizes was coded.

Allegiance analyses were carried out first for SPJ tools, and then, as a point of comparison, separately for actuarial tools. In the first analysis that focused on SPJ tools,

allegiance was coded as being present if any of the authors of a study also was a co-author of the original version of the measure being investigated ($k = 20$) or a co-author of an officially sanctioned translation of the measure ($k = 18$).[33] The majority of studies ($k = 68$) were completed by investigators who were not associated with the development or translation of an SPJ measure.

Analyses indicated the absence of an allegiance effect: Q_B ($df = 2$) = 1.09, $p = .58$ (see Table 1.11). Similarly, a significant effect was not observed when the two operationalizations of allegiance were collapsed into the same category ($k = 38$) and compared to studies in the non-allegiance category ($k = 68$).

Table 1.11. Allegiance

Allegiance Status	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
SPJ Measures					
Present, Original Version	20	2,453	.67 (.02)	.63 - .72	10.84
Present, Translated Version	18	1,852	.71 (.02)	.66 - .76	15.06
Absent	68	10,527	.68 (.01)	.66 - .70	64.70
Actuarial Measures					
Present	8	3,890	.69 (.02)	.64 - .73	1.75
Absent	60	10,627	.66 (.01)	.64 - .69	53.74

Note. k = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. For overall test for SPJ measures, $n = 14{,}832$; $k = 106$; $AUC_w = .68$ (SE = .01); 95% CI = .67 - .70 ; Q_B ($df = 2$) = 1.09, $p = .58$; Q_W ($df = 103$) = 90.60, $p = .80$; $Z = 72.95$, $p \leq .001$. For overall test for actuarial measures, $n = 14{,}517$; $k = 68$; $AUC_w = .67$ (SE = .01); 95% CI = .65 - .69 ; Q_B ($df = 1$) = .60, $p = .44$; Q_W ($df = 66$) = 55.49, $p = .83$; $Z = 69.33$, $p \leq .001$. All levels of the moderators have weighted mean effect sizes that differed significantly from chance, as indicated by a z-test ($p \leq .001$).

In light of the significant allegiance effect reported by Marcus et al. (in press) pertaining to three actuarial measures (Static-99, SORAG, and VRAG), exploratory analyses were undertaken to investigate whether an allegiance effect may be operating for the small (and nonrepresentative) grouping of actuarial measures coded in the present meta-analysis. Again, no statistically significant differences in the mean weighted effect sizes between samples for which

[33] See http://www.sfu.ca/psyc/faculty/hart/Guides.htm.

there was overlap between a study's author and the author of an actuarial measure under investigation ($AUC_w = .69$) versus samples for which no such overlap existed ($AUC_w = .66$; Q_B ($df = 1$) = .60, $p = .44$).

SPJ Tools: A Closer Look

Association between SPJ Risk Assessment Tools and Antisocial Behaviour

A detailed analysis of the effect sizes contributed by individual SPJ tools was undertaken. Tables 1.12 through 1.16 present results from analyses using the 'MetaES' macro across the various criterion categories (when effects were available for inclusion) for total score, summary risk rating, and scale scores of the five most researched SPJ measures: HCR-20, SVR-20, SARA, SAVRY, and ERASOR. Table 1.17 presents effect sizes available for additional SPJ measures (START, SCJ: Risk, HKT-30, EARL-20B, RSVP, and SORM,) for which a research base is only starting to accrue.

HCR-20

Table 1.12 presents aggregate results for 328 weighted AUC values for judgments of risk based on the final summary risk rating and HCR-20 total and scale scores across the various categories of antisocial behaviour. No significant levels of heterogeneity among effect sizes was observed for any HCR-20 index at any of the outcome categories. The largest mean weighted effect size ($AUC_w = .79$) was observed for summary risk ratings when the criterion was physical (+ sexual violence). In fact, for all outcome categories for which effects were available for meta-analytic analyses, the summary risk rating was larger in magnitude than any other HCR-20 index—including

Table 1.12. HCR-20 and Types of Antisocial Behaviour

Antisocial Index	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Summary Risk Rating					
Any Antisocial Behaviour	8	995	.70 (.03)	.64 - .77	5.81
Violent	3	331	.76 (.06)	.66 - .86	1.37
Physical (+ Sexual)	3	415	.79 (.05)	.69 - .88	1.44
Non-violent	3	347	.67 (.06)	.57 - .78	.03
Total Score^					
Any Antisocial Behaviour	51	6756	.69 (.01)	.66 - .71	43.68
Violent	14	1548	.73 (.03)	.68 - .78	7.16
Physical (+ Sexual)	30	4642	.67 (.02)	.64 - .70	33.57
Sexual	2	350	.46 (.08)	.31 - .61	1.97
Non-violent	16	1926	.70 (.02)	.66 - .74	12.82
Historical Scale					
Any Antisocial Behaviour	50	6302	.67 (.01)	.65 - .70	32.32
Violent	10	1022	.70 (.03)	.64 - .76	6.56
Physical (+ Sexual)	31	4329	.66 (.01)	.63 - .69	20.21
Physical (- Sexual)	2	226	.75 (.07)	.62 - .88	.68
Non-violent	15	1227	.64 (.03)	.58 - .69	9.32
Clinical Scale					
Any Antisocial Behaviour	46	5631	.65 (.02)	.61 - .68	61.37
Violent	8	685	.69 (.04)	.62 - .77	2.88
Physical (+ Sexual)	29	3903	.63 (.02)	.60 - .66	30.79
Physical (- Sexual)	2	226	.63 (.07)	.50 - .76	.05
Sexual	3	510	.69 (.06)	.57 - .81	3.60
Non-violent	16	1443	.68 (.03)	.63 - .73	7.82
Risk Management Scale					
Any Antisocial Behaviour	39	5167	.65 (.02)	.62 - .68	45.30
Violent	7	581	.71 (.04)	.63 - .79	1.81
Physical (+ Sexual)	23	3360	.63 (.02)	.59 - .67	28.15
Physical (- Sexual)	2	372	.62 (.05)	.52 - .72	.0001
Non-violent	12	1250	.68 (.03)	.62 - .73	4.00

Note. *k* = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. Physical (+ Sexual) = Physical Aggression Including Sexual Violence; Physical (- Sexual) = Physical Aggression Excluding Sexual Violence. IPV = Intimate Partner Violence. Levels of all variables have weighted mean effect sizes that differed significantly from chance, as indicated by a *z*-test ($p \leq .001$). ^ 14 disseminations were excluded from this analysis because only one or two of the three HRC-20 scales were used (NB: results changed minimally when they were included in analyses as contributing an ES for total score). For the summary risk rating, only one effect size was available for the physical (- sexual) violence category; none were presented for the sexual violence and IPV categories. For the Total Score, only one effect size was available for the physical (- sexual) violence category; no effect size was presented for the IPV category. For the Historical Scale, only one effect size was available for the sexual violence and IPV categories. For the Clinical Scale, no effect size was presented for the IPV category. For the Risk Management Scale, no effect sizes were available for the sexual violence and IPV categories.

the total score—with the exception of the nonviolent aggression category. For the summary risk rating, total score, and the Historical, Clinical, and Risk Management scales, the mean weighted AUCs (i.e., representing the weighted average effect size across, collapsed across all outcome categories presented in a study) ranged from .67 to .79, .46 - .73, .64 - .75, .63 - .69, and .62 - .71, respectively. It should be noted that analyses for some HCR-20 index/outcome category pairings were based on a small number of effect sizes, as noted in Table 1.12.

SVR-20

Table 1.13 presents aggregated results for 74 weighted effect sizes for various SVR-20 indices and outcomes. The largest SRR effect size (AUC_w = .70) corresponded to estimates of risk for sexual violence - the specific type of outcome that the SVR-20 was developed to assess. The same mean weighted effect size also was observed for estimates of risk based on the total score when the outcome was any antisocial behaviour. Of the three scales, effect sizes were smallest in magnitude for the Sexual Offences Scale. No statistically significant degree of heterogeneity was observed among any of the categories, although the Q for the Sexual Offences Scale and sexual violence approached significance (p = .06).

SARA

Table 1.14 presents aggregated results for 45 weighted effect sizes for various SARA indices and outcomes. As was the case for the HCR-20 and SVR-20, the largest mean SRR weighted effect size observed (AUC_w = .73) corresponded to estimates of risk based on summary risk ratings associated with the type of outcome for which the measure was developed to assess risk—intimate partner aggression. However, the largest AUC value (.79) was for violence predicted by the total SARA score; however, this mean value is based on only two effect sizes, and there was a statistically significant

Table 1.13. SVR-20 and Types of Antisocial Behaviour

Antisocial Index	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Summary Risk Rating					
Any Antisocial Behaviour	4	335	.69 (.05)	.59 - .79	2.08
Physical (- Sexual)	4	355	.66 (.05)	.56 - .77	2.31
Sexual	4	355	.70 (.07)	.57 - .84	4.94
Total Score					
Any Antisocial Behaviour	10	1124	.70 (.03)	.64 - .77	10.10
Physical (+ Sexual)	2	287	.60 (.08)	.45 - .75	1.68
Physical (- Sexual)	6	598	.65 (.04)	.57 - .73	4.11
Sexual	8	822	.61 (.05)	.52 - .70	11.82
Non-violent	3	366	.63 (.06)	.51 - .75	2.71
Psychosocial Adjustment Scale					
Any Antisocial Behaviour	4	394	.66 (.05)	.57 - .76	.39
Physical (+ Sexual)	2	220	.67 (.07)	.54 - .80	.04
Physical (- Sexual)	3	260	.67 (.06)	.55 - .80	.08
Sexual	3	260	.58 (.07)	.45 - .71	2.27
Sexual Offences Scale					
Any Antisocial Behaviour	4	394	.58 (.05)	.49 - .68	1.15
Physical (+ Sexual)	2	220	.56 (.07)	.43 - .69	.25
Physical (- Sexual)	3	260	.55 (.06)	.43 - .67	.09
Sexual	3	260	.61 (.11)	.40 - .82	5.69^
Future Plan Scale					
Any Antisocial Behaviour	3	343	.62 (.06)	.51 - .74	2.21
Physical (+ Sexual)	2	220	.57 (.07)	.44 - .70	.47
Physical (- Sexual)	2	209	.65 (.07)	.51 - .78	.18
Sexual	2	209	.69 (.08)	.52 - .85	1.46

Note. *k* = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. Physical (+ Sexual) = Physical Aggression Including Sexual Violence; Physical (- Sexual) = Physical Aggression Excluding Sexual Violence. IPV = Intimate Partner Violence. Levels of all variables have weighted mean effect sizes that differed significantly from chance, as indicated by a *z*-test ($p \leq .001$). No effect sizes were available for the following predictor/criterion pairings: SRR/Violent. SRR/Physical (+ Sexual), SRR/IPV, Total score/Violent, Total score/IPV, Psychosocial Adjustment Scale/Violent, Psychosocial Adjustment Scale/IPV, Sexual Offences Scale/Violent, Sexual Offences Scale/IPV, Future Plan Scale/Violent, Future Plan Scale/IPV. Only one effect size was available for the following predictor/criterion pairings: SRR/Non-violent, Psychosocial Adjustment Scale/Non-violent, Sexual Offences Scale/IPV, Future Plan Scale/Non-violent. ^ p = .06.

amount of variability present. Although predictive validity estimates for the total score and summary risk ratings were acceptable, relatively smaller effect sizes were observed for Parts 1 and 2, with Part 1 performing at chance level for the outcome of any antisocial behaviour. No

statistically significant degree of heterogeneity was observed among any of the categories, although the Q for Total score and 'violent' category approached significance ($p = .08$).

Table 1.14. SARA and Types of Antisocial Behaviour

Antisocial Index	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Summary Risk Rating					
Any Antisocial Behaviour	4	563	.66 (.06)	.55 - .78	5.06
IPV	4	563	.73 (.04)	.64 - .81	1.79
Total Score					
Any Antisocial Behaviour	9	3755	.67 (.03)	.60 - .73	26.34
Violent	2	261	.79 (.20)	.40 - 1.18	10.17**
IPV	8	3602	.63 (.02)	.60 - .66	3.99
Part 1					
Any Antisocial Behaviour	4	451	.50 (.07)	.50 - .75	5.53
Violent	2	261	.69 (.11)	.47 - .90	2.99^
IPV	3	298	.58 (.06)	.46 - .69	.31
Part 2					
Any Antisocial Behaviour	4	451	.59 (.05)	.60 - .69	1.07
Violent	2	261	.57 (.06)	.44 - .69	.07
IPV	3	298	.67 (.06)	.56 - .78	.38

Note. k = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. Physical (+ Sexual) = Physical Aggression Including Sexual Violence; Physical (- Sexual) = Physical Aggression Excluding Sexual Violence. IPV = Intimate Partner Violence. No effect sizes were available for any predictor with the following outcome categories: Physical (+ Sexual), Physical (- Sexual), and Sexual. Only one effect size was available for nonviolent aggression for all SARA indices, and for the SRR/Violent predictor/criterion pairing. Levels of all variables have weighted mean effect sizes that differed significantly from chance, as indicated by a z-test ($p \leq .001$). ^ $p = .08$.

SAVRY

Table 1.15 presents aggregated results for 168 weighted effect sizes for SAVRY indices and antisocial outcomes. As was the case for the other SPJ tools thus far, the largest mean weighted effect size observed ($AUC_w = .79$) corresponded to estimates of risk based on structured professional judgment and specifically for the type of outcome the measure was developed to assess—general violence. Total and scale scores also had strong predictive validity indices. Of note, the Protective Factors scale also performed well, with an AUC_w of .73 for the physical (+ sexual) violence category. No statistically significant degree of heterogeneity was observed

among any of the categories, although the Q for Historical scale approached significance ($p = .08$) for the physical (+ sexual) violence category.

Table 1.15. SAVRY and Types of Antisocial Behaviour

Antisocial Index	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Summary Risk Rating					
Any Antisocial Behaviour	10	1039	.69 (.03)	.62 - .75	10.40
Violent	4	377	.71 (.05)	.61 - .81	2.22
Physical (+ Sexual)	5	493	.79 (.05)	.69 - .89	5.15
Non-violent	6	658	.70 (.04)	.62 - .77	.94
Total Score					
Any Antisocial Behaviour	16	1481	.72 (.03)	.67 - .77	10.69
Violent	5	443	.73 (.05)	.64 - .82	.88
Physical (+ Sexual)	8	720	.75 (.04)	.67 - .82	7.77
Non-violent	8	798	.68 (.04)	.61 - .75	2.56
Historical Items					
Any Antisocial Behaviour	12	1141	.65 (.03)	.59 - .71	12.44
Violent	3	256	.64 (.06)	.52 - .76	.84
Physical (+ Sexual)	7	615	.66 (.05)	.55 – 77	11.36^
Non-violent	4	471	.65 (.05)	.56 - .74	.48
Social/Contextual Items					
Any Antisocial Behaviour	12	1141	.67 (.03)	.61 - .73	12.09
Violent	3	256	.65 (.06)	.52 - .77	2.12
Physical (+ Sexual)	7	615	.70 (.05)	.61 - .80	8.82
Non-violent	4	471	.64 (.05)	.55 - .73	1.14
Individual/Clinical Items					
Any Antisocial Behaviour	12	1141	.71 (.04)	.64 - .78	15.27
Violent	3	256	.72 (.08)	.55 - .87	3.31
Physical (+ Sexual)	7	615	.75 (.04)	.67 - .84	7.46
Non-violent	5	506	.68 (.04)	.60 - .77	1.61
Protective Items					
Any Antisocial Behaviour	12	1141	.68 (.04)	.61 - .75	17.19
Violent	3	256	.59 (.06)	.47 - .71	1.03
Physical (+ Sexual)	7	615	.73 (.06)	.62 - .85	12.82
Non-violent	5	506	.67 (.04)	.58 - .75	1.73

Note. k = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. Physical (+ Sexual) = Physical Aggression Including Sexual Violence; Physical (- Sexual) = Physical Aggression Excluding Sexual Violence. IPV = Intimate Partner Violence. For all SAVRY indices, only one effect size was available for the Physical (+ Sexual) and Sexual outcome categories, and no effects were available for the IPV category. Protective Items are reversed score and predict non-reoffending. ^ $p = .08$.

ERASOR

Table 1.16 presents aggregated results for 58 weighted effect sizes for ERASOR indices and antisocial outcomes. As is evident, none of the indices on the ERASOR consistently demonstrated acceptable levels of predictive accuracy. Of course, these findings are based on very few studies (two or three), and thus should be interpreted with prudence. Moreover, the performance of estimates of risk based on structured professional judgments could not be evaluated for the ERASOR given that only one weighted effect size for the summary risk rating was available (for five outcome categories).

Other SPJ Measures

Several additional promising measures developed within the SPJ model have been developed recently. Table 1.17 lists AUC values for six measures and any outcome categories that were available. Among the five measures for use with adults—RSVP, START, SCJ: Risk, HKT-30, and SORM—effect sizes for the measures' total scores ranged from .72 - .63 for any antisocial behaviour. Notably, ample range in magnitude of effects was observed across outcome categories and scale scores for the HKT-30 and SCJ: Risk.

Estimates of risk based on summary risk ratings currently are not available for START, SCJ: Risk, or SORM. For two of the tools for which summary risk ratings were available, AUC values based on these judgments were larger than judgments based on the numeric score (e.g., for the RSVP, .73 vs. 63; for the EARL-20B, .73 vs. .56 and .68). This was not true for the HKT-30, however, where some variability between effect sizes based on summary risk ratings and numeric scores was observed (see Table 1.17).

Table 1.16. ERASOR and Types of Antisocial Behaviour

Antisocial Index	k	n	AUC_w (SE)	95%CI_w	Q
Total Score					
Any Antisocial Behaviour	3	425	.59 (.06)	.47 - .71	3.08
Physical (- Sexual)	2	297	.63 (.06)	.51 - .74	.09
Sexual	3	425	.61 (.07)	.48 - .75	3.80
Non-violent	2	297	.45 (.06)	.33 - .57	1.11
Sexual Interests					
Any Antisocial Behaviour	3	425	.59 (.05)	.49 - .69	2.11
Physical (- Sexual)	2	297	.55 (.06)	.43 - .66	.22
Sexual	3	425	.59 (.08)	.45 - .74	4.35
Non-violent	2	297	.51 (.06)	.40 - .62	.09
Historical Sexual					
Any Antisocial Behaviour	3	425	.53 (.05)	.43 - .63	2.29
Physical (- Sexual)	2	297	.57 (.06)	.45 - .68	.61
Sexual	3	425	.54 (.08)	.39 - .69	4.53
Psychosocial Functioning					
Any Antisocial Behaviour	3	425	.60 (.06)	.47 - .72	3.07
Physical (- Sexual)	2	297	.67 (.06)	.56 - .79	.21
Sexual	3	425	.53 (.07)	.39 - .67	3.99
Non-violent	2	297	.55 (.06)	.43 - .66	.92
Family Functioning					
Any Antisocial Behaviour	3	425	.56 (.05)	.46 - .65	1.49
Physical (- Sexual)	2	297	.52 (.06)	.41 - .63	.04
Sexual	3	425	.62 (.05)	.52 - .71	.25
Non-violent	2	297	.54 (.06)	.42 - .65	.79
Treatment					
Any Antisocial Behaviour	3	425	.50 (.05)	.42 - .60	.36
Physical (- Sexual)	2	297	.53 (.06)	.41 - .64	.02
Sexual	3	425	.49 (.05)	.39 - .58	.51
Non-violent	2	297	.48 (.06)	.36 - .59	.41

Note. *k* = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. Physical (+ Sexual) = Physical Aggression Including Sexual Violence; Physical (- Sexual) = Physical Aggression Excluding Sexual Violence. IPV = Intimate Partner Violence. No effect sizes were available for any of the ERASOR indices for Violent and IPV outcome categories, or for Historical scale/Non-violent. Additionally, only one effect size was available across ERASOR indices for the Physical (+ sexual) outcome category. Protective Items are reversed score and predict non-reoffending.

Table 1.17. Other SPJ Tools Included in the Meta-analysis

Risk Assessment Measure	Antisocial Index AUC						
	Any	Violent	Physical (+ Sexual)	Physical (- Sexual)	Sexual	IPV	Non-violent
RSVP							
Total Score		.67			.63		
Summary Risk Rating					.73; .73; .73		
START Total Score	.70		.70				.67; .72
SCJ: Risk							
Total Score	.68	.66					.71
Historical Scale	.56	.54					.58
Clinical Scale^	.72	.68					.77
Risk Management Scale^	.66	.63					.58
Suicide Scale	.61	.62					.62
Vulnerability Scale	.57	.59					.63
Escape Scale	.53	.53					.52
HKT-30							
Total Score	.61; .72		.61	.60; .72	.55		
Summary Risk Rating	.59; .73		.59	.55; .73	.61		
Historical Scale	.62; .67		.62	.65; .67	.45		
Clinical Scale	.55; .69		.55	.52; .69	.58		
Future Scale	.65; .68		.65	.63; .68	.58		
SORM	.67	.71					
EARL-20B							
Total Score	.56; .68						
Summary Risk Rating	.73						

Note. Values in the column for 'any' outcome represent weighted mean weighted effect sizes. Values in other columns are single, unweighted AUCs. Physical (+ Sexual) = Physical Aggression Including Sexual Violence; Physical (- Sexual) = Physical Aggression Excluding Sexual Violence. IPV = Intimate Partner Violence. ^Same items as HCR-20 Scale with the same name.

Association between Different Risk Assessment Approaches and Antisocial Behaviour Outcomes

When available, effect sizes for indices (total and scale scores) from non-SPJ methods of assessing risk were coded as well. Such indices comprised various actuarial measures, measures

from the PCL family, and unstructured clinical judgment. These effects in no way are intended to be presented as a representative sample of effects sizes for actuarial (and psychopathy) measures, but, rather, are included merely for purposes of comparison. The availability of these effect sizes provides the opportunity for direct, within study comparisons between SPJ and non-SPJ methods of assessing risk for violence to others.

Whereas descriptive statistics presented earlier in the results section represented the 949 effect sizes for any SPJ index, what follows is a description of the total 1665 effect sizes coded (including the 949 for SPJ) for the meta-analysis. Following that overview, analyses will be presented using effect sizes culled from this larger group of $k = 1665$ to explore the comparative performance of the different assessment approaches for making judgments about risk.

Central Tendency and Distribution

Table 1.18 summarizes the central tendency and distribution of raw (not weighted) and weighted AUC values as a function of risk assessment approach. In the first of the three layers in the Table, findings based on all 1665 effect sizes for the two primary approaches to risk assessment under investigation (SPJ and actuarial) are presented. For comparison purposes, findings for risk assessments based on the PCL family of measures and unstructured clinical judgment additionally are presented. The second layer of the table also presents findings that are not statistically independent; these values represent the mean weighted effect sizes across nonredundant effects for each measure within each sample. For example, if a study presented effect sizes for the LSI-R total score as well as for individual LSI-R scales for both violent and nonviolent outcomes, the AUC coded would be the average of the total scores across the mutually exclusive outcome categories. Finally, the third layer of the Table presents findings for effect sizes at the level of one effect size per risk assessment approach per sample. For example, the values for the SPJ approach represent average within study judgments of risk based on total scores and summary risk ratings from all SPJ tools used in a particular study (k = 104). Similarly,

values for the actuarial approach represent the average within study judgments of risk based on total scores from all actuarial measures used in a particular study (k = 45).

Table 1.18. Association between Risk Assessment Approach and Antisocial Behaviour: Distribution of AUCs

Assessment Approach	*k*	*n*	Not weighted				Weighted		
			Mean (SE)	*SD*	Mdn	Q1 – Q3	AUC_w (SE)	95%CI_w	*Q*
All AUCs									
SPJ									
All AUCs	949	116,858	.66 (.01)	.11	.66	.59-.73	.65 (.003)	.65-.66	1198.86***
Summary risk rating	83	8,633	.68 (.01)	.11	.68	.59-.75	.68 (.01)	.65-.70	105.79*
Total scores	282	36,868	.68 (.01)	.11	.68	.62-.76	.68 (.01)	.66-.69	336.10**
Scale scores	584	71,357	.65 (.01)	.11	.64	.57-.71	.64 (.004)	.63-.65	718.43***
Actuarial									
All AUCs	334	62,948	.61 (.01)	.10	.62	.55-.67	.62 (.01)	.61-.63	560.27***
Total scores	253	51,563	.62 (.01)	.11	.62	.55-.69	.62 (.01)	.61-.64	509.90***
Scale scores	81	11,385	.60 (.01)	.06	.60	.55-.65	.60 (.01)	.58-.62	43.13
PCL									
All AUCs	357	53,405	.65 (.01)	.10	.65	.60-.71	.66 (.01)	.65-.67	389.45
Total scores	153	22,413	.66 (.01)	.09	.67	.60-.73	.67 (.01)	.66-.69	132.31
Scale scores	204	30,992	.64 (.01)	.11	.65	.58-.69	.64 (.01)	.63-.66	249.35**
UCJ	18	1,655	.60 (.01)	.05	.61	.57-.63	.59 (.02)	.55-.64	4.89
One AUC per Measure per Sample									
SPJ									
All	143	19,109	.69 (.01)	.10	.70	.63-.76	.68 (.01)	.67-.70	142.63
Final risk judgment	32	3,595	.69 (.02)	.10	.71	.63-.74	.69 (.02)	.65-.72	27.01
Total scores	110	15,514	.70 (.01)	.10	.69	.62-.76	.68 (.01)	.67-.70	115.58
Actuarial Total Scores	70	15,113	.67 (.01)	.08	.67	.61-.72	.67 (.01)	.65-.68	70.85
PCL Total Scores	55	7,516	.66 (.01)	.07	.66	.60-.72	.67 (.01)	.64-.69	28.92
UCJ	6	699	.58 (.03)	.08	.60	.52-.64	.58 (.04)	.50-.65	3.16
One AUC per Risk Assessment Approach per Sample									
SPJ ^	104	15,077	.70 (.01)	.10	.70	.63-.76	.68 (.01)	.67-.70	114.17
Actuarial$^{\Omega}$	45	9,037	.67 (.01)	.08	.68	.64-.73	.67 (.01)	.65-.70	37.90
PCL	55	7,516	.66 (.01)	.07	.66	.60-.71	.66 (.01)	.64-.69	26.19
UCJ	6	699	.58 (.03)	.08	.60	.52-.64	.58 (.04)	.50-.65	3.16

Notes. Criterion is any antisocial behaviour. *k* = number of effect sizes; SE = standard error of the mean; Mdn = Median; Q1 = first quartile; Q3 = third quartile. AUC_w(SE)= mean weighted effect size (AUC) and standard error of the weighted mean effect size; 95% CI_w = 95% confidence interval of mean weighted effect size. SPJ = Structured Professional Judgment. Final risk judgment refers to structured, categorical ratings of risk (i.e., low, moderate, or high). PCL = Psychopathy Checklist family of measures. UCJ = Unstructured Clinical Judgment. Effect sizes (AUCs) in the first and second horizontal layers are dependent. ^ Represents the average within sample judgments of risk based on total scores and summary risk ratings from all SPJ tools used in a particular sample. $^{\Omega}$ Represents the average within study judgments of risk based on total scores from all actuarial measures used in a study. $p \geq .05$. $^{**}p \geq .01$. $^{***}p \geq .001$. Seven of the 1665 effect sizes represent numerical (i.e., actuarially based) SPJ risk categories and were excluded from these analyses.

A stem-and-leaf plot is presented in Figure 1.1 to assist in visualizing the shape of the distribution of all effect sizes coded. As was the case when the subsample of effects for SPJ measures was examined, the distribution of all effect sizes also closely approximated a normal curve.

Figure 1.4. Stem-and-Leaf Plot of 1665 Raw AUC Values

Frequency	Stem	& Leaf
18.00	Extremes	(=<.38)
7.00	3 .	9&
24.00	4 .	14&
73.00	4 .	56778899
161.00	5 .	00011122233334444444
224.00	5 .	55556666677777888888999999
318.00	6 .	000000011111112222222223333333334444444
328.00	6 .	55555555666666667777777777888888888889999
251.00	7 .	00000011111112222233333344444
132.00	7 .	5555666677788999
67.00	8 .	0012334
33.00	8 .	5678&
5.00	9 .	&
24.00	Extremes	(>=.92)

Notes. & = fractional leaves; Stem width = .100; Each leaf = 8 cases.

Examining findings at the one-per-sample level, the last layer in Table 1.18 indicates that the overall magnitude of the weighted effect size for the SPJ approach is slightly larger than that for the actuarial approach, although the difference is small (.70 vs. .67). The PCL measures performed as well as the actuarial approach, and all approaches were substantially more useful in terms of predictive validity than unstructured clinical judgment (although it should be noted that the 95% confidence intervals for unstructured clinical judgment overlapped at least minimally with the confidence intervals for the other approaches).

Actuarial Tools: A Closer Look

Association between Actuarial Risk Assessment Tools and Antisocial Behaviour

Table 1.19 presents mean weighted effect sizes for actuarial measures as a across the seven outcome categories. Values ranged from AUC_w = .69 for physical (including sexual) violence to AUC_w = .61 for nonviolent aggression. The magnitude of these effect sizes is somewhat smaller than that of effect sizes for the SPJ approach presented in Table 1.3.

Table 1.19. Actuarially-Based Risk Assessment Approach and Antisocial Behaviour

Antisocial Index	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Any Antisocial Behaviour	45	8330	.67 (.01)	.65-.69	34.26
Violent	9	1201	.68 (.03)	.62-.74	3.63
Physical (+ Sexual)	19	2906	.69 (.02)	.65-.72	11.78
Physical (- Sexual)	9	1056	.64 (.03)	.58-.70	5.92
Sexual	15	1724	.64 (.02)	.59-.68	8.76
Intimate Partner Aggression	4	2160	.62 (.02)	.58-.66	2.46
Non-violent	13	1767	.61 (.04)	.54-.68	23.43

Note. *k* = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. Levels of all variables have weighted mean effect sizes that differed significantly from chance, as indicated by a *z*-test ($p \leq .001$). Data represent independent effect sizes (i.e., one effect size per sample), which is the aggregate of within study average of risk assessments based on an actuarial measure.

There were 24 different measures than contributed effect sizes to these analyses: Danger Assessment Scale (DAS; Campbell, 1986, 1995); Domestic Violence Risk Appraisal Guide (DVRAG; Hilton, Harris, Rice, Houghton, & Eke, 2008); Domestic Violence Screening Instrument (DVSI; Williams & Houghton, 2004); Statistical Information on Recidivism Scale (SIR; Nuffield, 1982); Juvenile Sex Offender Assessment Protocol -II (J-SOAP-II; Prentky & Righthand, 2003); Juvenile Sexual Offense Recidivism Risk Assessment Tool -II (J-SORRAT-II; Epperson, Ralston, Fowers, & DeWitt, 2005); Kingston Screening Instrument for Domestic

Violence (KSID; Gelles & Tolman, 1998); Lifestyle Criminality Screening Form (LCSF; Walters, White, & Denney, 1991); Ontario Domestic Assault Risk Assessment (ODARA; Hilton et al., 2004); Offender Group Reconviction Scale (OGRS; Copas & Marshall, 1998); Preliminary Scheme (PS; Hartvig, Alfarnes, Skjønberg, Moger, & Østberg); Rapid Risk Assessment for Offense Recidivism (RRASOR; Hanson, 1997); Violence Risk Appraisal Guide (VRAG; Harris et al., 1993; Quinsey et al., 2006); Sex Offender Risk Appraisal Guide (SORAG; Quinsey et al., 1998); Static-99 (Hanson & Thornton, 1999); Risk Matrix 2000 (RM 2000; Thornton et al., 2003); Short Dynamic Risk Scale (SDRS; Quinsey, Book, & Skilling, 2004); Salient Factor Score (SFS; Hoffman, 1983); Static-2002 (Hanson & Thornton, 2003); Violent Offender Risk Assessment Scale (VORAS; Howells, Watt, Hall, & Baldwin, 1997); Violence Risk Scale, second edition (VRS-2; Wong & Gordon, 2000); Youth Level of Service/Case Management Inventory (YLS/CMI; Hoge & Andrews, 2002); Level of Service Inventory–Revised (LSI-R) (Andrews & Bonta, 1995); Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta & Wormith, 2004).

VRAG

To examine more closely findings for the actuarial measure for which the most effect sizes were coded in the present meta-analysis, effect sizes across the various outcome categories for the VRAG are presented in Table 1.20. The largest mean weighted effect sizes were observed for nonviolent aggression and physical violence that did not include sexual violence (AUC_w = .74). Values for other outcome categories ranged from AUC_w = .59 (sexual violence) to AUC_w = .70 (any antisocial behaviour).

Table 1.20. VRAG and Types of Antisocial Behaviour

Antisocial Index	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Any Antisocial Behaviour	19	4148	.70 (.02)	.67-.73	8.83
Violent	4	508	.68 (.04)	.59-.77	.94
Physical (+ Sexual)	11	2534	.71 (.02)	.67-.75	10.05
Physical (- Sexual)	3	220	.74 (.07)	.61-.88	1.26
Sexual	3	220	.59 (.08)	.43-.75	2.78
IPV	2	737	.68 (.04)	.61-.75	.50
Non-violent	3	1214	.74 (.03)	.68-.79	.09

Note. *k* = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. Levels of all variables have weighted mean effect sizes that differed significantly from chance, as indicated by a *z*-test ($p \leq .001$).

Assessment of Psychopathic Personality Disorder: A Closer Look

Association between PCL Tools and Antisocial Behaviour

Mean weighted estimates of predictive validity across outcome categories were computed for the PCL-R and PCL:SV (6 effect sizes for PCL:YV total scores were excluded from these analyses to restrict the predictor items to a comparable group). As noted earlier, these data represent only a small subsample of all available research on the PCL measures. Effects are presented here only as a point of comparison to effects for the SPJ approach. There were 49 effect sizes available for PCL-R ($k = 29$) and PCL:SV ($k = 20$) total scores. Of those, 28 also presented findings for Part/Factor scores. Although some studies reported findings for the more recent three-factor (Cookie & Michie, 2001) and four-facet (Hare, 2003) factor structures, the majority of findings available in the literature were for the original two-factor structure.

For PCL-R/SV total score, mean weighted effect sizes varied from $AUC_w = .70$ for physical violence that did not include sexual violence ($k = 6$) to $AUC_w = .60$ for sexual violence ($k = 8$). Effect sizes varied more widely in magnitude across outcome categories for the Factors/Parts. For both Factor/Part 1 and 2, the largest effect size was observed for the 'violent'

outcome category (AUC_w = .70 and .71, respectively). Compared to effect sizes for Factor/Part 2, effects were smaller for Factor/Part 1, where the smallest effect size observed was at chance level for physical (excluding sexual) violence (AUC_w = .51) (see Table 1.21).

Table 1.21. PCL-R/PCL:SV and Types of Antisocial Behaviour

Antisocial Index	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
Total Score					
Any Antisocial Behaviour	49	6747	.66 (.01)	.64-.69	23.62
Violent	15	1445	.67 (.03)	.62-.72	5.29
Physical (+ Sexual)	24	3773	.66 (.02)	.63-.69	23.71
Physical (- Sexual)	6	540	.70 (.04)	.62-.79	1.37
Sexual	8	794	.60 (.04)	.53-.67	6.03
IPV	3	839	.65 (.04)	.59-.72	1.26
Non-violent	14	1974	.65 (.02)	.61-.70	14.46
Factor/Part 1					
Any Antisocial Behaviour	28	3729	.62 (.02)	.59-.65	24.58
Violent	5	438	.70 (.01)	.59-.81	5.22
Physical (+ Sexual)	19	2992	.61 (.02)	.57-.65	15.42
Physical (- Sexual)	2	135	.51 (.09)	.34-.68	.38
Sexual	3	351	.55 (.16)	.24-.86	14.13
Non-violent	10	1334	.61 (.02)	.56-.65	7.78
Factor/Part 2					
Any Antisocial Behaviour	28	3729	.68 (.02)	.65-.71	14.53
Violent	5	438	.71 (.05)	.62-.81	1.22
Physical (+ Sexual)	19	2992	.67 (.02)	.64-.71	14.87
Physical (- Sexual)	2	135	.68 (.09)	.51-.84	.46
Sexual	3	351	.55 (.05)	.44-.65	.52
Non-violent	11	1159	.68 (.02)	.64-.72	7.54

Note. *k* = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size. Physical (+ Sexual) = Physical Aggression Including Sexual Violence; Physical (- Sexual) = Physical Aggression Excluding Sexual Violence. IPV = Intimate Partner Violence. Levels of all variables have weighted mean effect sizes that differed significantly from chance, as indicated by a *z*-test ($p \leq .001$). Only one effect size was available for the Factors/Parts for IPV.

Direct Comparisons of SPJ and Actuarial Tools

The performance of SPJ tools when used as intended for clinical decision making purposes (i.e., summary risk ratings) was compared to the performance of actuarial tools. The ideal approach to investigating the relative performance of these two approaches involves making

direct comparisons between the AUC values reported within each study to gauge the frequency with which comparisons are statistically significantly different, and in what direction. This analytic approach was untenable because most studies did not report the standard error that accompanied the AUC value, which is required to test the difference between two correlated AUC values (see formula 3, Hanley & McNeil, 1983). As an alternative, a meta-analytic analogue to the ANOVA was completed using Lipsey and Wilson's (2001) MetaF macro. Using only those studies in which an SPJ tool and actuarial measure were examined in the same sample, an ANOVA was run for each of four categories of violence: any antisocial behaviour, violent, physical (+ sexual), and physical (-sexual). The performance of the SPJ tools when numeric judgments are used also was compared to the performance of actuarial tools.

As is evident in Table 1.22, regardless of whether the summary risk rating or the numeric rating was used, the mean weighted effect sizes for comparisons between the SPJ and actuarial approaches were moderate in size and virtually identical for all comparisons, although values were slightly larger for the SPJ approach in some cases. None of the models for the eight comparisons was statistically significant. Model values for each comparison were as follows: SRR/Actuarial for 'any antisocial behaviour', Q_B (df = 1) = .16, p = .69; SPJ total/Actuarial for 'any antisocial behaviour', Q_B (df = 1) < .001, p = .99; SRR/Actuarial for 'violent', Q_B (df = 1) < .001, p = .98; SPJ total/Actuarial for 'violent', Q_B (df = 1) = .41, p = .52; SRR/Actuarial for 'physical + sexual violence', Q_B (df = 1) = .46, p = .50; SPJ total/Actuarial for 'physical + sexual violence', Q_B (df = 1) = .12, p = .73; SRR/Actuarial for 'physical-sexual violence', Q_B (df = 1) = .07, p = .79; and SPJ total/Actuarial for 'physical-sexual violence', Q_B (df = 1) = .04, p = .83.

Table 1.22. Direct Comparisons between Approaches across Violence Categories

Risk Comparison	*k*	*n*	AUC_w (SE)	95%CI_w	*Q*
		Any Antisocial Behaviour			
SRR vs. Actuarial	10	1031			
SRR			.65 (.03)	.58-.72	9.88
Actuarial			.63 (.03)	.56-.70	7.05
SPJ Total vs. Actuarial	44	8634			
SPJ Total			.67 (.01)	.65-.70	38.43
Actuarial			.67 (.01)	.65-.70	42.12
		Violent			
SRR vs. Actuarial	2	161			
SRR			.61 (.08)	.46-.76	1.06
Actuarial			.61 (.08)	.46-.76	.90
SPJ Total vs. Actuarial	9	1190			
SPJ			.71 (.03)	.65-.76	3.43
Actuarial			.68 (.03)	.62-.74	3.63
		Physical (+ Sexual)			
SRR vs. Actuarial	2	265			
SRR			.73 (.06)	.61-.85	1.58
Actuarial			.67 (.06)	.55-.79	.51
SPJ Total vs. Actuarial	20	3765			
SPJ Total			.69 (.02)	.65-.72	18.18
Actuarial			.68 (.02)	.64-.71	17.95
		Physical (- Sexual)			
SRR vs. Actuarial	6	601			
SRR			.61 (.04)	.53-.69	4.09
Actuarial			.62 (.04)	.54-.71	4.60
SPJ Total vs. Actuarial	10	1138			
SPJ Total			.66 (.03)	.60-.71	6.04
Actuarial			.65 (.03)	.59-.71	5.93

Note. *k* = number of effect sizes; AUC_w = mean weighted effect size (AUC); SE = standard error of the weighted mean effect; 95% CI_w = 95% confidence interval of mean weighted effect size; SRR = summary risk rating; SPJ = structured professional judgment. Physical (+ Sexual) = Physical Aggression Including Sexual Violence; Physical (- Sexual) = Physical Aggression Excluding Sexual Violence. Levels of all variables have weighted mean effect sizes that differed significantly from chance, as indicated by a *z*-test ($p \leq .001$). *n* listed is the number of people in the sample assessed on both measures.

DISCUSSION

Within a fairly short amount of time—only a few decades—the field of what has come to be known as violence risk assessment has transformed dramatically (Hanson, 2005; Monahan, 1996; Monahan & Steadman, 2001; Steadman, 2000). The small number of studies conducted on this issue by the early 1980s suggested that clinicians' ability to assess risk accurately was unremarkable. Reinterpretation of the literature conducted by the late 1980s using different analytic approaches indicated comparatively higher levels of predictive accuracy (Mossman, 1994), and the development of new risk assessment technologies has resulted in even higher indices of accuracy reported in the empirical literature.

Literally hundreds of measures have been developed to assess risk for future violence, and meaningful research has informed development of several formal risk tools that are psychometrically sound (see, e.g., Borum, 1996; Douglas, Cox, et al., 1999; Douglas, Ogloff, Nicholls, & Grant, 1999; Douglas & Webster, 1999; Hart, 1998; Monahan et al., 2001b; Quinsey et al., 2006; Steadman et al., 2000). Formal risk tools tend to be based either on an actuarial or clinical approach to decision making. The SPJ framework for assessing risk for violence, a clinically based approach in which structure is imposed on the decision making process in specific ways, was developed to overcome the perceived shortcomings of the unstructured clinical approach (e.g., unreliability, subjectivity) and the actuarial approach (e.g., failure to include relevant risk factors, difficulties with generalization, insufficient ability to inform violence management and prevention efforts). Although many individual studies have examined these tools' performance, and meta-analyses of subsets of the SPJ literature are available (Campbell et al., 2007; Hanson et al., 2007; Nikolova et al., 2006; Reeves et al., in preparation), no comprehensive meta-analytic evaluation of the SPJ model—without restrictions imposed on inclusion criteria—have been conducted.

This dissertation offers such an evaluation, and its results are informative for at least two reasons. First, various SPJ violence risk assessment tools are used internationally in many clinical settings, and it is critical to understand the extent and nature of the empirical foundation for this practice, including whether there exist any factors that may moderate the magnitude of tools' predictive accuracy. Second, given the serious consequences that may result from violence risk assessments both to the individual being evaluated and to society, it is vital to direct empirical scrutiny on recommendations made by some researchers that the actuarial approach be the only approach used when conducting risk assessments (Quinsey et al., 2006). To the extent that the field of psychology is in a position to offer services (risk assessments) using particular empirically validated technologies (SPJ risk assessment tools) that potentially may protect society and/or shield against infringement upon examinee's civil liberties, it seems irresponsible to not continue research into (and promulgation of) such technologies.

Summary of and Commentary on Main Findings

Overall Predictive Accuracy of the SPJ Model

Taken together, the independent effect sizes contributed by the 104 disseminations indicate that a solid empirical foundation underlies the SPJ model of violence risk assessment. The mean weighted AUC of the SPJ approach, when any antisocial behaviour was the outcome under study, was .68 (the non-weighted value was nearly identical; AUC = .70). Translating to more traditional indices of association, an AUC value of .68 is roughly equal to a *d* value of .67 and a point biserial correlation of .32 (Rice & Harris, 2005).

Most manuals of SPJ tools explicitly recommend that evaluators communicate findings of their risk assessment by means of a summary risk rating (typically, but not necessarily, using the low/moderate/high trichotomy), rather than summing the item scores and reporting a numeric score. In terms of predictive accuracy, in the present investigation, the observed weighted AUC

values for summary risk ratings were larger than the values associated with total scores across five of the seven outcome categories examined. Although for some outcomes (any antisocial behaviour, non-violent behaviour, violent, nonsexual physical violence), the differences between effect sizes for the total score and summary risk rating were negligible, the differences were larger for intimate partner aggression and physical/sexual violence. Indices of predictive accuracy for the SPJ model were highest for the violent and physical/sexual violence categories, with the largest weighted AUC (.77) observed for the summary risk rating for the physical/sexual violence group.

These results provide support for the utility of the SPJ model in general, and also indicate that professionals are able to make structured clinical judgments that perform as well as or better than numeric use of SPJ tools. Of course, the predictive accuracy of the summary risk ratings and numeric applications would be expected to be, at a minimum, comparable because, in general, the more risk factors that are present for a given individual, the higher his or her risk would be. At the same time, however, it would be inappropriate to interpret the comparable performance of the two indices as reason to discontinue use of summary risk ratings; among other reasons, such ratings are critical for situations in which one or a few risk factors are deemed to be disproportionately important to an examinee's risk for violence; this is information that would be lost were item scores simply to be summed. The importance of the summary risk rating to the SPJ approach and recommendations for its use will be discussed in further detail in the Implications for Clinical Practice section.

Predictive Accuracy of Individual SPJ Tools

Despite support for the SPJ decision-making approach in the aggregate, the present results also point to a deficit in empirical foundation for certain SPJ tools. With the exception of the HCR-20 and SAVRY, no SPJ tools had over a dozen evaluations of its predicative validity. Of course, no magical threshold exists that indicates the minimum number of studies required to

demonstrate a sufficient foundation for predictive validity, but given the serious consequences that can occur as a result of violence risk assessments, it is important for the foundation to be robust.

In general, more research has evaluated tools for use with adults and adolescents than children. The literature search identified only two SPJ tools for use with children: the EARL-20B for boys and EARL-21G for girls, both of which were developed at the Child Development Institute in Toronto, Ontario. For each tool, only two predictive validity evaluations were identified. Research findings for the EARL-21G were not included in this meta-analysis because, for one study (Levene et al., 2004), an effect size could not be calculated based on the data provided,[34] and for the other study (Yuile, 2008), it was unclear whether and if so how an effect size could be calculated.[35] Although additional research is underway to evaluate these promising tools, given the scarcity of the empirical findings at present, empirical support for the clinical use of these tools is limited at present.

Two tools were identified for use with adolescents—the SAVRY and the ERASOR. Although a relatively substantial amount of predictive validity research has been completed with the SAVRY, only three such evaluations were identified for the ERASOR. On the basis of the small amount of empirical investigation of the ERASOR to date, coupled with the finding that the overall weighted AUC for the outcome it was intended to assess—sexual violence—was .61, further study need be made of the ERASOR to investigate the appropriateness of its use in clinical practice.

[34] When contacted, the authors generously offered to make such findings available in the near future.

[35] Yuile (2008) completed an evaluation of a multisystemic intervention in which the EARL-21G was administered to 162 girls. The author presented results of mixture model analyses in which she examined the association between child and family risk factors and girls' subsequent antisocial behaviours. Attempts to clarify the nature of the analytic presentation in order to calculate an effect size with the author were unsuccessful. A subsequent attempt to clarify the analyses using the expertise of a statistical consultant (personal communication, R. Koopman, July 6, 2008) also were unsuccessful.

In contrast, results of the performance of the SAVRY were very supportive of its continued use. Predictive accuracy indices for the SAVRY across various outcome categories for both the summary risk rating and total score were at least on par with the overall average observed across all SPJ tools. Findings especially offered support for the utility of the summary risk rating; the largest weighted AUC for the SAVRY was for the summary risk rating, for physical/sexual violence (.79). Among the SAVRY's scales, the strongest performance was for the Protective Items scale (AUC_w for physical/sexual violence = .73), which highlights the importance of examining strengths of an adolescent, as well as risks. AUC values for the Social/Contextual and Individual/Clinical Items were slightly higher than values for the Historical Items. This may reflect the fluctuating nature of developmental shifts that occur during adolescence. Static, historical factors clearly are relevant to forecasting antisocial outcomes, but failure to consider the individual and contextual factors that currently are relevant for a youth would seem to dilute the predictive accuracy of the risk assessment. This point is discussed in more detail in a section below.

Regarding measures for use with adults, the literature search identified 8 SPJ tools that had at least one evaluation of its predictive validity. The most researched SPJ tool was the HCR-20, which is not surprising given that it was the first tool to be developed according to the SPJ model for violence risk assessment. As was the case for the SAVRY, the largest weighted AUC for the HCR-20 was for the summary risk rating, for physical/sexual violence (.79). All three of the tool's scales performed well for predicting violence, but there were variations in accuracy across the categories of outcomes.

An important finding regarding the performance of the HCR-20 of which to take note is its lack of accuracy for assessing risk for sexual violence. The weighted AUC associated with the total score was .46 (although this is based on only two effect sizes), whereas the largest values were observed for general/physical violence. This is not surprising; the risk factors identified for

inclusion on the HCR-20 were selected on the basis of their demonstrated empirical association with general, as opposed to sexual violence, or any other specific form of violence. The finding highlights the importance of selecting a tool that is best suited to address the purpose of the particular clinical task.

Two adult SPJ tools were identified that assess specialized forms of violence: the SVR-20 for sexual violence and the SARA for intimate partner violence. Compared to the HCR-20, relatively less research has been conducted with these tools, and even fewer investigations have examined the performance of their summary risk ratings (vs. their total scores). Findings for both tools again demonstrated that, when the performance of the summary risk rating is evaluated, predictive accuracy is maximized when there is a match between the type of violence the tool was developed to assess and the outcome under consideration. For the SVR-20, the largest SRR AUC was for sexual violence (.70). This finding is based on only four effect sizes, however. Whereas the largest values for two of the three SVR-20 scales (Sexual Offences Scale and Future Plan Scale) also were for sexual violence, this was not the case for the total score, where the largest value was for the general category of any antisocial behaviour. The Psychosocial Adjustment and Future Plan Scales had somewhat higher indices of accuracy than the Sexual Offences Scale.

For the SARA, the largest SRR AUC was for intimate partner violence (.73). Similar to the SVR-20, this finding also is based on only four effect sizes. The SARA total score was relatively more accurate for assessing risk for general violence (.79) than for intimate partner violence (.63) specifically. Predictive validity indices were much higher when assessments were based on all the information in the SARA, rather than on one of its two Parts, with forecasts being near chance level for some outcome categories for Parts 1 and 2.

Finally, there also were several measures for use with adults identified, but which have been researched only minimally. Additional evaluations of the performance of these promising tools—START, RSVP, SCJ: Risk, HKT-30, and SORM—are required.

Comparative Performance of SPJ and non-SPJ Approaches

In addition to offering empirical justification for the use of the SPJ decision-making model, the present findings also essentially void criticisms (Quinsey et al., 2006) that clinically-based estimates of violence risk, even when structured and anchored by research findings, are inappropriate for use in clinical practice. Analysis of the predictive validity for the non-SPJ tools[36] indicated that, in general, they performed as well as the SPJ tools. The weighted AUC for all actuarial tools was .67 (compared to .68 for SPJ tools). When the VRAG, which was the actuarial tool with the most effect sizes coded in this project, was examined on its own, the weighted AUC for any antisocial outcome was .70, which was comparable to the performance of SPJ tools. This value also is highly similar to findings reported by Blair et al in (in press), who found that the overall predictive validity of the VRAG based on 21 studies was r = .32, which would translate into an AUC value of approximately .69 (Rice & Harris, 2005). The performance of the PCL-R total score, as well as both Factors, also was comparable to the actuarial measures for assessing risk for violence.

The most informative analysis regarding the comparative performance of SPJ and actuarial tools was the examination of effect sizes from studies in which both types of tools were used with the same sample. Without exception, across the four types of outcome studied, estimates of predictive validity based on SPJ tools were as large as or larger than estimates based on actuarial tools. The largest discrepancy in AUC values was observed for the physical/sexual violence category for which the weighted AUC values for the summary risk rating and actuarial tools were .73 and .67, respectively. It should be noted, however, that this finding is based on only two studies. Results of findings from studies in which direct comparisons were made indicate that estimates of predictive validity, which is only one aspect upon which the usefulness

[36] The reader is reminded that the sample of non-SPJ tools is not representative of the literature, as non-SPJ tools were included in this analysis only if they appeared in a study in which an SPJ tool was examined; thus, a large section of the literature in which the predictive validity of non-SPJ tools was examined is excluded here.

of a violence risk assessment scheme should be evaluated, are similar across both types of decision making models.

Moderator Analyses

Six variables of relevance to clinical practice were examined for their potential to moderate predictive accuracy. Irrespective of the gender composition of the sample, the age of participants, the country in which data were collected, the setting in which the risk assessment occurred, the setting in which violence occurred, or whether an interview was used to complete the risk assessment, no statistically significant variability in the magnitude of effect sizes was observed. A trend toward significance, however, was observed for gender, with assessments tending to be more accurate for women. Two variables that are more directly relevant to conducting research also were examined. Although the magnitude of the effect size did not differ as a function of allegiance, the research design did exert a moderating influence. Taken together, findings from the moderator analyses provide strong support for the generalizability of the SPJ model. Results for each variable will be discussed in turn below.

Gender

Researchers have reported that although rates of violence appear to be similar among psychiatric samples of women and men (Nicholls, 2001; Nicholls et al., 2004), clinicians' (unstructured) predictions of violence tend to underestimate risk potential among women (see, e.g., Coontz et al., 1994; Elbogen et al., 2001; McNeil & Binder, 1995; Skeem, Schubert, et al., 2005). The present results suggest that when risk factors are selected and considered in a structured manner, the gender bias may be neutralized. Although moderator analyses for gender were not statistically significant, there was a trend towards significance ($p = .06$), but in the opposite direction to what would be expected based on research examining unstructured clinical assessments. That is, there was a trend for predictive validity to be higher when SPJ tools were

used with women (AUC_w = .78) compared to men (AUC_w = .68). It should be noted that only 10 effect sizes were available for analysis for the female samples, versus 60 for the male samples. The imbalance in number of effect sizes indicates that additional research examining the performance of SPJ tools with women is needed.

Age

Only two samples included in the meta-analysis examined the predictive accuracy of an SPJ tool with children (both of which used the EARL-20B), and thus discussion of the moderator analyses for age will focus on results of studies for adults (18 years of age and older) and adolescents (between 13 and 17 years of age). Age was not a statistically significant moderator of predictive accuracy. However, somewhat surprisingly, the predictive accuracy for the SPJ model was slightly higher for samples of juveniles (AUC_w = .75) than adults (AUC_w = .67).

Although no hypothesis was made regarding the impact of age on predictive accuracy, this finding is somewhat surprising in light of the unique challenges to assessing risk for violence among juveniles presented by the nature of the developmental shifts that occur during adolescence (Cauffman & Steinberg, 2000a, 200b; Grisso, 1998; Steinberg & Cauffman, 1996, 1999). To a degree, engaging in some form of antisocial behaviour is normative among adolescents, and the clinical assessment challenge therefore entails efforts to minimize false positives. However, most adolescents who exhibit some antisocial behaviour cease to do so as they progress through subsequent developmental stages. For example, at least half of all children who exhibit serious and persistent antisocial behaviour before age 10 are not violent as adolescents (Patterson, Forgatch, Voerger, & Stoolmiller, 1998). Similarly, the majority of seriously antisocial adolescents do not continue on to be antisocial adults (Moffitt & Caspi, 2001; Moffit, Caspi, Dickson, Silva, & Stanton, 1996).

Risk assessment with juveniles also necessitates consideration of developmental stage and social context (Mulvey, 2005). More concretely, risk factors can have different implications

for risk likelihood as a function of developmental stage (see, e.g., Odgers, Vincent, & Corrado, 2002). For example, drinking alcohol before age 10 is a significant risk factor, but drinking at age 16, which could be interpreted as an indication of experimentation and is normative for this developmental stage, is not necessarily a risk factor. The present results suggest that the individualized nature of assessing risk using the SPJ approach works well with adolescents in which the relative importance of risk factors may be prone to change as they transition through developmental stages.

Nationality

Almost all of the SPJ tools identified in the present review were developed in Canada.[37] As such, it is important to consider the extent to which predictive validity may differ when SPJ tools are used in other countries. Another impetus for studying nationality was that, as reviewed earlier, this variable has been found to moderate the predictive accuracy of psychopathy as operationalized by the PCL-R or PCL: SV in prisons or jails (Guy et al., 2005). In contrast to previous findings, the present results indicate that SPJ tools are robust against any influence originating from the country in which the data were collected, and, as such, offer support for the generalizability of the SPJ model.

Clinical Setting

Whether participants were drawn from forensic psychiatric, civil psychiatric, or correctional settings did not impact the magnitude of SPJ tools' predictive accuracy. Although details regarding the nature and extent of mental disorder across participants in forensic psychiatric, civil psychiatric, and correctional settings was not examined in the present study, this finding is consistent with a large scale meta-analysis by Bonta and colleagues, who reported that

[37] Tools not developed in Canada include: SCJ: Risk, which was developed in the U.K., but is based on the HCR-20; HKT-30, which was developed in the Netherlands, but also is based on the HCR-20; and SORM, which was developed in Sweden.

the major predictors of recidivism were the same for mentally disordered and nondisordered offenders. These results may be understood in terms of reflecting the importance of the flexibility that is a defining feature of the SPJ approach, and which contributes to the model's generalizability.

Violence Location

The finding that predictive accuracy did not differ as a function of whether violence occurred in the community versus in a hospital, jail, or prison also may be understood to reflect the flexibility of the SPJ approach. Certain contextual factors would be expected to affect the likelihood of the occurrence of violence, such as security levels in an institution that would restrict opportunity for engaging in violence against others. The finding that predictive validity of risk assessments made using the SPJ approach was constant across this variable provides further support for the generalizability of this decision-making model.

Source of Information Used to Complete Risk Assessment

Unexpectedly, whether an evaluator interviewed the individual or only relied on file information to complete the risk assessment did not moderate the magnitude of predictive validity. This was surprising because certain variables may be expected to be assessed more accurately in person, such active symptoms of mental illness. However, under some circumstances, highly detailed file information may provide sufficient information for assessing the presence (and absence) of risk factors. Indeed, evaluators in Canada typically have access to files that are remarkably detailed and include past psychiatric reports.

The present finding notwithstanding, it would be inappropriate to complete a risk assessment for clinical purposes without having attempted to interview the individual being evaluated. The code of ethics adhered to by members of the American Psychological Association (APA, 2002) notes that, except when not warranted, "psychologists provide opinions of the

psychological characteristics of individuals only after they have conducted an examination of the individuals adequate to support their statements or conclusions" (APA, 2002, p. 13). This practice also would be consistent with the spirit of informed consent presented in the Canadian Code of Ethics for Psychologists (2000 p. 10), which directs psychologists to "Seek as full and active participation as possible from others in decisions that affect them, respecting and integrating as much as possible their opinions and wishes." Of course, there are circumstances under which risk assessments may be completed without input from the evaluee (e.g., security level classification decisions upon reception at a correctional or psychiatric facility) or when the evaluee refuses to participate (e.g., forensic assessments ordered by a court; assessments to aid release decision-making among tribunals or parole boards).

Study Design

Retrospective designs were associated with larger effect sizes than studies that employed pseudo- or true-prospective designs. It is noteworthy that, despite the convenience and comparative ease of conducting retrospective evaluations, only a small minority ($k = 16$) of effect sizes were contributed by studies with retrospective designs. This is important because, fundamentally, the question for which evaluators are asked to provide an opinion is risk for *future* violence. One reason that retrospective designs yielded larger indices of accuracy may be related to criterion contamination, though this is difficult to evaluate because of the 16 samples with retrospective designs, there were 10 cases in which it could not be discerned whether raters of the risk tool were blind to outcome. Of the remaining 6 samples, raters were noted specifically not to be blind to outcome in 5 instances. In only one sample were raters described as being blind to outcome. Based on the 6 samples for which definitive information is known, it would have been possible that having knowledge regarding a participants' history of antisocial behaviour may have contributed to assignment of theoretically consistent ratings (i.e., lower ratings for cases with less dense offending histories, and vice versa).

Allegiance

Given the manifestly different approaches to violence risk assessment espoused by developers of actuarial and SPJ tools, combined with claims by both groups regarding the appropriateness of their respective methods, it is important to examine the extent to which researchers' allegiance to a tool may affect the magnitude of predictive accuracy observed. In the psychotherapy literature, researchers' therapy allegiance is a strong predictor of which psychosocial treatment is observed to be superior at the individual study level (Luborsky et al., 1999). In the violence risk assessment literature, Blair et al. (in press) reported evidence for an allegiance effect in a meta-analysis that focused on the predictive validity of three actuarial measures (VRAG, SORAG, and Static-99). Contrary to these findings, no evidence of an allegiance effect was observed either for SPJ or actuarial tools in the present meta-analysis. Lack of an allegiance effect in the present study is consistent with the finding that predictive accuracy was not moderated by whether a study had been published in a peer-reviewed outlet, which is relevant because an allegiance effect may be observed as a result of the file-drawer problem (Rosenthal, 1979).

Another reason why allegiance effects may exist is that measure authors' studies incorporate particular study design characteristics associated with larger effects. In the present meta-analysis, however, no variables except study design were found to moderate the magnitude of predictive accuracy. Of course, not all potentially relevant study characteristics were examined.

Allegiance effects also may be observed because researchers who developed the instrument may be expected to be relatively more proficient in the administration and application of the tool, which could impact the validity of a risk assessment. Arguably, this factor may be more important for SPJ tools than for actuarial tools, which ostensibly have strict and explicit

scoring guidelines. The fact that an allegiance effect was not observed among studies of SPJ tools offers further support for the validity of the SPJ model.

Implications for Clinical Practice

Cumulatively, the present findings provide support for the use of: (a) the SPJ decision-making model when conducting violence risk assessments; (b) summary risk ratings to communicate level of risk posed; and (c) SPJ tools developed to assess specialized, rather than broad, types of violence for assessments of risk for particular categories of violence.

Results further indicate that the SPJ model is generalizable across a variety of contexts and circumstances. This characteristic is what supports the development of SPJ tools designed to meet local demands. Because manuals of SPJ tools indicate that they should be considered as guidelines or aide-memoires, adapting existing guidelines for use in different jurisdictions and disparate settings is not problematic. For example, the SCJ: Risk, used in the U.K., and the HKT: 30, used in the Netherlands, are adaptations of the HCR-20. Authors of both tools incorporated additional factors to be responsive to certain legal and contextual factors relevant in their jurisdictions. Given the way in which SPJ guidelines are developed, they are much more amenable to application across settings and jurisdictions that actuarial measures, whose content and structure cannot be altered.

Of particular importance to clinical practice, the present findings clearly contradict claims that the actuarial approach is superior to any and all clinically-based approaches to assessing risk for violence. Effect sizes were virtually indistinguishable in analyses that compared the overall predictive accuracy of SPJ and actuarial models across all studies, as well as in analyses in which comparisons were limited only to those studies in which instruments from both models were examined using the same sample. This is consistent with recent meta-analytic reviews (Campbell et al., 2007; Hanson et al., 2007; cf. Hanson & Morton-Bourgon, 2004).

In light of the similarity in predictive accuracy between different approaches, what other factors should a professional consider when selecting a risk assessment tool? In the early stages of development in the violence risk assessment field, the focus of clinical work and research was on one-time, dichotomous predictions of whether violence would occur; over time, significant shifts have occurred such that prevention and management of violence are now seen as paramount. Although most actuarial assessments may offer useful information in one sense, such as informing baseline recommendations regarding the anticipated intensity of intervention, monitoring, and/or supervision efforts that would be required, they typically are unable to inform subsequent prevention and management efforts (see, e.g., Heilbrun, Nezu, Keeney, Chung, Wasserman, 1998) because they tend to include static risk factors, with few or no dynamic risk factors. An important exception, discussed below, is the LSI family of tools. To that extent, and given the comparable performance in terms of predictive validity across the models, SPJ tools (and actuarial tools that include criminogenic risk factors such as the LSI tools) confer distinct and important advantages critical to assessing and managing risk of violence.

In addition to the SPJ model, another well researched approach to violence risk assessment that also has a strong empirical foundation is the Risk-Need-Responsivity (RNR) model (see, e.g., Andrews & Bonta, 2006; Andrews, Bonta, & Wormith, 2006; Andrews & Dowden, 2006; Andrews, Bonta, & Hoge, 1990; Gendreau & Andrews, 1990). The performance of this model was examined in detail in the meta-analysis by Campbell et al. (2007) given their inclusion of the LSI family of instruments. The RNR model presents an approach for risk assessment and classification of individuals (the model initially was developed for use with offenders) for treatment. The RNR model is based on a general social learning model of deviance (Andrews et al., 1990). Three principles define the model and are instructive regarding how intervention should be delivered. First, the risk principle suggests that offenders at higher risk for reoffending will benefit maximally when given higher levels and intensity of intervention,

whereas lower risk offenders should be offered relatively less (or no) intervention. Second, the need principle advises that interventions should target criminogenic needs, which refer to the dynamic attributes of an offender that, when changed, are associated with changes in the probability of recidivism. Finally, according to the responsivity principle, mode of intervention should be matched to certain characteristics about the individual, including learning style, level of motivation, and contextual/interpersonal circumstances.

In addition to evidence that tools developed according to the SPJ and RNR models offer comparable estimates of predictive validity (Campbell et al., 2007), the models are similar in other respects, which will be described below. Perhaps the most important way in which the models differ is the approach taken to assess risk. The risk tools associated with the RNR model are the LSI family of instruments,[38] which are actuarial in nature. As described by the founders of the model, "The items were selected on the basis of prior evidence of their association with recidivism (and cross-validated with the original LSI construction sample). Additionally, LS risk/need entails mechanical scoring with evidence-based contingency tables linking score categories with measures of outcome" (Andrews et al., in press, p. xx). As such, the way in which items are selected for SPJ and RNR tools are similar, but differences are apparent in the ways in which the risk factors are combined, as well as the way in which the final risk estimate is generated. That is, within the RNR approach, there are explit rules that direct the way in which the items are combined and the way in which the final risk estimate is generated. In contrast, within the SPJ approach, the evaluator's discretion is valued for making both of these decisions.

[38] For an overview, see Andrews et al. (in press). Briefly, the revised version of the first measure developed according to this model, the LSI-R, is a 54-item quantitative measure developed to inform assessment of risk and need with offenders. It provides a structured assessment of 10 criminogenic domains: Criminal history, Education/Employment, Financial, Family/Marital, Accommodation, Leisure/Recreation, Companions, Alcohol/Drug Problem, Alcohol/Drug Problem, Emotional/Personal, and Attitudes/Orientation. Total scores can be translated into risk bands to determine risk of reoffending.

Although there clearly are differences between the models, there also are many similarities. First, there is a relatively high degree of content overlap between the tools developed according to each model to assess general risk among adults (i.e., the LSI-R and the HCR-20). Second, both models emphasize the importance of assessing dynamic risk factors. Third, the models both view professional discretion as valuable, which is the focus of the fourth principle of the RNR model. This principle, which has been interpreted mostly as pertaining to making treatment recommendations, asserts that clinical judgment should override the other three principles should the particular circumstances warrant. Finally, and of most relevance to this discussion, although both models assess risk for reoffending, they additionally focus on risk management and prevention.

Given the similarities between the models, research has investigated potential ways in which risk tools from the SPJ and RNR models could be used concurrently to augment predictive accuracy and provide unique risk management information. Guy et al. (2007) examined the correspondence between the scales/domains of the HCR-20 and LSI-R among 226 men admitted to a forensic psychiatric hospital in Germany. Substantial correspondence between the HCR-20's scales and the LSI-R's risk/need domains suggested several areas that were assessed by one measure, but not tapped by the other. Additional risk and need areas assessed by the LSI-R that could contribute unique information about present clinical and future risk management variables on the HCR-20 included increased consideration of leisure and recreation activities and use of alcohol and drugs.

Implementation Issues

Although the use of SPJ tools is known to occur in a many settings both domestically and abroad, violence risk assessment instruments in general have not been incorporated into standard clinical practice, despite the strong empirical foundation for a number of such tools. Challenges to translating scientific knowledge about risk assessment into clinical practice are multifaceted.

Elbogen (2002) discussed how a lack of knowledge by professionals regarding violence risk assessment tools limits their use in practice. Another obstacle involves deficits in resources required to administer the instruments (e.g., Steadman et al., 2000). Appropriate and well-founded concerns regarding the validity of certain tools within various populations also have been cited as posing a barrier to use (e.g., Otto, 2000).

Packer (2001) surveyed the State Forensic Director of each of the 50 United States and the District of Columbia to assess the extent to which various violence risk assessment tools are used in public sector forensic practice, and reasons for lack of use. The response rate was 64%. Reasons for not using a formal tool included the instruments not being validated with the particular population, instruments being too time consuming, lack of staff training in use of an instrument, lack of awareness of such instruments, cost, and unavailability of criminal background data (an essential element to using most tools). Packer concluded that several of the impediments to use of structured instruments can be easily ameliorated, such as need for staff training and educating administrators about such instruments, but that concerns regarding validity issues require additional discussion and research to address. To that end, the next section discusses the implications for research raised by the findings of this dissertation.

Implications for Research

Taken together, results of this dissertation offer support for the predictive validity of the SPJ model. Given the nature of meta-analytic techniques, however, potentially important variability in the validity estimates was obscured, which has implications for further research. Although almost all moderator analyses were not significant, it is possible that a strategic analysis at the individual study level—focusing on the disseminations that contributed the largest and smallest effect sizes for a given SPJ tool—may provide essential clues regarding sample or design features that may moderate the association between risk assessment and recidivism.

Through the process of aggregating research conducted on the SPJ model, certain gaps in the field's knowledge regarding decision-making using this approach in particular and about risk assessment in general became apparent. Most notably, at least four areas are worthy of additional research: (a) evaluation of the primary purpose of the SPJ model, which includes considering issues related to summary risk ratings, technology transfer, and dynamic predictive validity as it is related to prevention efforts; (b) demographic characteristics, including gender, age, and race/ethnicity; (c) facets of violence; (d) and protective factors.

First, and perhaps most importantly, further research is warranted to evaluate the primary purpose of the SPJ model. That is, the usefulness of summary risk ratings to inform risk management and treatment deserves further empirical inquiry. Most studies that have evaluated the predictive validity of an SPJ tool have done so using (or at least only reporting data for) the numerical scores, rather than the summary risk ratings. As noted by others (e.g., Douglas, 2001; Heilbrun et al., in press), according to the test standards jointly derived and adopted in 1999 by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), the empirical evaluation of psychological measures should correspond to their intended clinical use(s). Additional research investigating not only the performance of summary risk ratings, but also their incremental validity relative to other types of judgments (i.e., numeric judgments from SPJ tools, judgments from actuarial instruments) (e.g., Edens, Skeem, & Douglas., 2006; Douglas et al., 2005) is needed.

Another aspect of the SPJ model greatly in need of evaluation is the process and impact of implementation, which refers to a specified set of activities intended to put into practice an activity or program (see Fixsen, Naoom, Blase, Friedman, & Wallace, 2005), of structured risk tools in correctional and psychiatric settings. Evaluations of intervention efforts can include analysis of intervention-level outcomes and implementation-level outcomes (Fixsenet al., 2005).

Intervention-level outcomes would include indicators such as changes in recidivism rates, changes in dispositions and placements (e.g., in cases of release decision-making, whether an individual, because of his or her risk level, is placed in a secure setting or community placement), and changes in service usage and completion following recommendations to participate in particular interventions. Implementation-level outcomes include indicators such as perceptions of staff and administrator regarding the feasibility and usefulness of the risk assessment tool, changes in staff knowledge regarding risk factors, changes in staff attitudes regarding rehabilitation and punitiveness, and changes in rates of risk level classification.

Though this technology transfer issue was discussed above as it pertains to implications for clinical practice, a discussion of the implications for research are warranted as well. Limited empirical study of SPJ tools has been undertaken in this area that is so critical to the translation of science into practice.[39] Evaluation of the implementation of the START in Canada was reported by Crocker and colleagues (no date). Studies are planned to examine the implementation of the SCJ: Risk in the U.K. (personal communication, C. Allen, August 5, 2008) and the SAVRY in the United States (personal communication, G. Vincent, August 20, 2008). Other evaluative components of implementation effort include focus group disseminations, which have been completed for some SPJ tools, such as the START (Crocker et al., n.d.).

The final aspect to be discussed here related to investigating the primary purpose of SPJ approach is evaluation of the model's goal of violence prevention (e.g., see Hart, 1998), which has not been the subject of much research. As noted by Douglas and Kropp (2002), one approach to validate this goal of the SPJ approach would involve evaluating whether the use of SPJ tools in concert with appropriate risk management strategies would reduce the subsequent base rate of violence to a larger degree than in the absence of such an intervention, a topic which overlaps

[39] There is a larger empirical base regarding the implementation of actuarial risk assessment tools (see, e.g., Flores, Lowenkamp, Holsinger, & Latessa, 2006; Haas, Hamilton, & Hanley, 2006; Young, Moline, Farrell, & Bierie, 2006).

with issues addressed above regarding evaluating implementation efforts. Although the importance of matching appropriate types and doses of interventions to risk level has been investigated within the context of the treatment for psychopathy (Skeem, Monahan, & Mulvey, 2002) and the RNR model (see, generally, Andrews & Bonta, 2006; Bonta & Andrews, 2007), no outcome data regarding the preventive goal of the SPJ model have been published.[40] In part, this would involve further empirical evaluation of dynamic predictive validity. As noted above, inclusion of dynamic risk factors is critical for identifying targets for intervention that should be associated with reductions in risk level and actual recidivism. However, most research on dynamic risk factors comprises one-time snapshots of the association between scores on scales that comprise dynamic risk factors and outcome. As explicated clearly by Douglas and Skeem (2005), more research is required to evaluate whether changes in risk factors over time are associated with changes in recidivism (hopefully with both changes occurring in the same direction).

Moving on to the second general area in need of further research, most samples identified for inclusion in the present review consisted of few women or children. In addition, data on ethnic and racial composition were limited. Regarding gender, it is recommended that future research include additional examinations of female samples. Moreover, when feasible, it would be helpful for researchers to report main findings for women and men separately. Regarding age, very few samples examined children. Although it is not surprising that most research has focused on populations in which antisocial behaviour is more likely to occur, additional research is required to evaluate the performance of the EARL-20B and -21G. Pertaining to race and ethnicity,

[40] However, an investigation currently underway by G. Vincent is examining the effect of implementing the SAVRY in "treatment" sites and withholding in "control" sites, and aims to investigate the preventive goal of the SPJ model. Kropp and colleagues also have begun to evaluate the "preventive utility" of incorporating a screening version of the SARA into practice among police officers in Sweden.

researchers should—at a minimum—report these data in disseminations. Ideally, sample descriptions would include some level of detail.

Moving now to the third general area in need of further research, though violence is known to be multifaceted, little investigation has been undertaken regarding critical aspects of violence, such as severity, imminence, duration and frequency (Dvoskin & Heilbrun, 2001; Hart, 1998, 2001; Heilbrun, 1997; Mulvey & Lidz, 1995; cf. Bengston & & Långström, 2007, who investigated these aspects for actuarial and unguided clinical judgments of risk for sex offenders). Douglas and Ogloff (2003a) offered the first empirical investigation of whether specific judgments (using the SPJ model) regarding particular facets of violence could be made reliably and whether the judgments could offer incremental validity relative to the more traditional types of judgments using the numerical or summary risk ratings. Although their findings suggested that more specific judgments about particular facets of violence were not made with the same reliability and validity as either the total scores or the structured clinical risk ratings, the study offers an excellent start to a topic area in need of further empirical investigation.

Finally, the importance of considering protective factors when conducting a violence risk assessment is clear (e.g., Rogers, 2000). Results of the present meta-analysis from the group of studies that investigated the SAVRY support the importance (in terms of predictive accuracy) of evaluating protective factors, and encourage the utility of future research in this area. Although the construct of resiliency, and the putatively protective factors that define resiliency, has been well studied in the field of developmental psychopathology (see, e.g., Cicchetti & Rogosch, 2002), the application of this knowledge to the field of violence risk assessment would benefit from additional empirical consideration. It is clear that "risk and protective factors (are not) simply each other's opposites" (Durlak, 1998, p. 518), and existing risk assessment schemes do in fact recognize protective factors as variables that, when present, reduce the risk of violence compared to when they are not present. For example, referencing the Protective Factors scale, the

authors of the SAVRY (2003, p. 9) noted: "While the absence of a risk factor may, in some sense, be considered "protective," and used accordingly in a risk appraisal, the factors contained in this section of the SAVRY are all positive protective factors (those that are notable for their presence, not their absence), as opposed to negative protective factors (those notable for the absence of a risk factor)."

In the adult risk assessment literature, however, considerably more attention has been focused on investigating pathology rather than areas of competence, and consequently protective factors traditionally have not been explicitly highlighted in risk assessment schemes (although clearly would be considered when formulating a risk management plan). For example, although the HCR-20 lists only 20 risk factors, in the companion guide (Douglas, Webster, Hart, Eaves, & Ogloff, 2001), chapters address how to capitalize on or shape individuals' weaknesses into strengths. Two recently developed SPJ schemes include an explicit focus on strengths. The SAPROF, discussed earlier, consists of 17 dynamic factors presented as being protective in nature. The authors describe the intended use of the SAPROF as being used "in combination with the HCR-20 or related structured risk assessment instruments." Although there appear to be some putative protective factors that are distinct from risk factors, many of the SAPROF's items are highly similar in content to the risk factors presented on the HCR-20. Similarly, the START presents a series of items for which the evaluator offers separate ratings regarding the extent to which each item represents a strength and risk for the individual. Although this approach is conceptually similar to what an evaluator using any SPJ scheme that uses the typical 0, 1, or 2 scale is invited to do when considering the extent to which there is evidence "for" or "against" an item, the START is distinguished for its initiative to offer explicit direction for "assessors to consider that a client can have both risks and strengths simultaneously" (Webster, Nicholls, et al., 2006, p. 28). Future research, especially in the adult literature, should build upon these efforts by identifying protective factors, learning more about at what point—for those risk and protective

factors that appear to exist on the same continuum—the pendulum swings from risk to protection, and studying how best protective factors may be integrated into existing risk assessment enterprises.

Limitations

The quality of any meta-analysis is dependent on, among other factors, the thoroughness of the literature search to identify studies that meet inclusion criteria. The present meta-analysis is no exception. Despite comprehensive efforts to identify research presented in peer-reviewed publications, in non-peer reviewed sources, and completed research that had not yet been disseminated at all, it is certain that not all existing and codeable studies were included in the present analyses. Titles of some studies that were identified as being potentially relevant could not be located despite efforts to do so; it would have been desirable to review these studies to ascertain if they met inclusion criteria (efforts are continuing to locate these studies). One study (Ogloff & Daffern, 2006) that presented data on the HCR-20's Clinical scale was brought to the attention of the author subsequent to the completion of analyses that should have been included but was omitted unintentionally.[41] As such, although the collection of studies included in the present analyses may be regarded as reflecting those identified through a comprehensive search effort, it is acknowledged that not all relevant studies were included.

Although this meta-analysis was able to examine empirically the comparative performance of actuarial and SPJ tools investigated using the same sample, analyses at the individual study level could not be completed because studies typically did not present the

[41] Ogloff and Daffern (2006) reported on the predictive validity of the C scale of the HCR-20 and two non-SPJ tools: the Broset Violence Checklist (BVC; Almvik, Woods, & Rasmussen, 2000) and the Dynamic Appraisal of Situational Aggression (DASA; Ogloff & Daffern, 2006). The AUC for the HCR-20 C scale was .73 and the AUC for the BVC was .83. The 7-item DASA consists of the items that demonstrated the largest association with physical aggression: two items from the HCR-20 (Negative Attitudes; Impulsivity); two items from the BVC (Irritability; Verbal threats), and three items from the authors' previous research (sensitive to perceived provocation; easily angered when requests are denied; unwillingness to follow directions). The AUC for these 7 items was 0.82.

requisite statistical information (individual standard errors associated with AUC values) for such analyses. As is the case for any meta-analysis, the analyses that may be completed are limited by the information presented in the primary sources. One solution to this obstacle would have involved contacting study authors to request that the analyses be recomputed and for the standard errors to be provided. However, this was deemed to be too demanding of a request to impose on researchers. Another way in which the "actuarial vs. clinical" debate issue could have been investigated in the present project would have involved examining results of incremental validity analyses presented in individual studies (when available), although the extent to which results would be appropriate for comparison across studies would depend on the similarity of the variables included in the incremental validity models.

Finally, this project is limited in the sense that there are many additional analyses that could have been completed but were not because they were beyond the scope of the primary research questions. Future research questions that may be investigated using the present dataset include whether there may be additional variables that moderate the association between SPJ judgments and violence, such as the length of time during which participants were at risk and whether the follow-up period was the same or varied among participants (e.g., Hanson & Morton-Bourgon, 2007); the base rate of detected violence and how many sources were used to identify the occurrence of violence (e.g., Buchanan, 1997; Douglas & Ogloff, 2003c; Mulvey, Shaw, & Lidz, 1994); and whether the risk assessment was completed for clinical ("real world") or research purposes.

Conclusion

The appropriateness of conducting violence risk assessments using clinically-based models of decision-making has been debated for decades. The SPJ approach was developed with the intention of ameliorating the deficits presented by both unstructured clinical and actuarial

approaches. Since the first set of SPJ guidelines was made available, hundreds of studies have been conducted that, taken together, provide empirical evidence that violence risk assessments completed according to the SPJ framework are both reliable and valid. This dissertation provides empirical support for the predictive validity of the SPJ model in assessing risk of violence to others. Results further support the generalizability of the SPJ approach across varied population and contextual parameters. The present research supports the continued use of certain SPJ tools for which there is evidence of sufficient predictive validity, but also indicates that additional investigations of less well researched SPJ tools would be extremely worthwhile. Finally, based on results of all retrievable research conducted to date on the SPJ model, this dissertation offers strong evidence that claims regarding the superiority of the actuarial approach for assessing risk for future violence are baseless.

REFERENCES

* References marked with an asterisk are included in the meta-analysis.

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341-382.

Airy, G. B. (1861). *On the algebraic and numerical theory of errors of observations and the combination of observations.* London: Macmillan.

*Allen, C., & Howells, K. (2008). *The implementation and evaluation of a structured professional judgment risk assessment tool within a high secure forensic hospital.* Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

Almvik, R., Woods, P., & Rasmussen, K. (2000). The Bröset Violence Checklist (BVC): sensitivity, specificity and inter-rater reliability. *Journal of Interpersonal Violence, 15*, 284-1296.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *The standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*, 1060–1073.

Americans with Disabilities Act. (1990). 42 U.S. Code Chapter 126.

*Andres-Pueyo, A. (2008). *Risk assessment for severe intimate partner violence using the Spanish adaptation of the SARA*. Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

*Andres-Pueyo, A., Perez, M., & Redondo-Illescas, S. (2008). *Recidivism and risk for sexual violence assessed by SVR-20 in a Spanish sample of treated sexual offenders.* Poster presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

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

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Cincinnati, OH: Anderson Publishing Company.

Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology, 50*, 88-100.

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. (in press). The Level of Service (LS) assessment of adults and older adolescents. In R. K. Otto, & K. S. Douglas (Eds.), *Handbook of violence risk assessment*. Oxford, United Kingdom: Routledge/Taylor & Francis.

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

Arnhoff, F. N. (1960). Some aspects of clinical judgment. *Journal of Clinical Psychology, 16*, 123-128.

*Augimeri, L. K., Enebrink, P., & Walsh, M. (in press). Gender-specific childhood assessment tools: Early Assessment Risk Lists for Boys (EARL-20B) and Girls (EARL-21G). In R. K. Otto, & K. S. Douglas (Eds.), *Handbook of violence risk assessment*. Oxford, United Kingdom: Routledge/Taylor & Francis.

Augimeri, L. K., Jiang, D., Koegl, C. J., & Carey, J. (2006). *Differential effects of the Under 12 Outreach Project (ORP) associated with client risk and treatment intensity*. Report submitted to the Provincial Centre of Excellence for Child and Youth Mental Health, Children's Hospital of Eastern Ontario (CHEO), Ottawa, Ontario, Canada.

Augimeri, L. K., Koegl, C. J., Ferrante, P., & Slater, N. (2006). Why and how: Conducting effective clinical risk assessments with children with conduct problems. *Canada's Children*, 24-27.

Augimeri, L. K., Koegl, C. J., Levene, K. S., & Webster, C. D. (2005). Early assessment risk lists for boys and girls. In T. Grisso, G. Vincent & D. Seagrave (Eds.), *Mental health screening and assessment in juvenile justice* (pp. 295-310). New York: Guilford Press.

Augimeri, L. K., Koegl, C. J., Webster, C. D., & Levene, K. S. (2001). *Early Assessment Risk Lists for Boys: Version 2*. Toronto, Ontario, Canada: Earlscourt Child and Family Centre.

Barbaree, H. E., & Seto, M. C. (1998). *The ongoing follow-up of sex offenders treated at the Warkworth Sexual Behaviour Clinic*. Toronto, Ontario, Canada: Centre for Addiction and Mental Health.

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

Barefoot v. Estelle, 463 U.S. 880 (1983)

*Bartel, P. A., Forth, A. E., & Borum, R. (n.d.). *The Development and Concurrent Validity of the Structured Assessment of Violence Risk in Youth (SAVRY)*. Unpublished manuscript.

*Bartel, P. A. & Regan, K. (2003, July). *The use of the Structured Assessment of Violence Risk in Youth (SAVRY) with adolescent females*. Presented at the Psychology and Law International, Interdisciplinary Conference, Edinburgh, Scotland.

Belfrage, H. (1998). Implementing the HCR-20 scheme for risk assessment in a forensic psychiatric hospital: Integrating research and clinical practice. *Journal of Forensic Psychiatry, 9,* 328-338.

Belfrage, H., & Douglas, K. S. (2002). Treatment effects on forensic psychiatric patients measured with the HCR-20 violence risk assessment scheme. *International Journal of Forensic Mental Health, 1,* 25-36.

Belfrage, H., & Strand, S. (in press). Validation of the Guide for Stalking Assessment and Management (SAM) in law enforcement: A prospective study of 153 cases of stalking in two Swedish police counties. *International Journal of Police Science and Management.*

Belfrage, H., & Strand, S. (2007, July). *Structured spousal violence risk assessment using the B-SAFER: Combining risk factors and victim vulnerability factors.* Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Montreal, Québec, Canada.

*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.

Bengston, S., & Långström, N. (2007). Unguided clinical and actuarial assessment of re-offending risk: A direct comparison with sex offenders in Denamrk. *Sex Abuse, 19,* 135-153.

Bill C-55: An Act to amend the Criminal Code (high risk offenders), the Corrections and Conditional Release Act, the Criminal Records Act, the Prisons and Reformatories Act and the Department of the Solicitor General Act. (1997). S. C., 1997, c. 17.

Blair, P. R., Marcus, D. K., & Boccaccini, M. T. (in press). Is there an allegiance effect for assessment instruments? Actuarial risk assessment as an exemplar. *Clinical Psychology: Research and Practice.*

Bloom, H., Eisen, R.S., Pollock, M., & Webster, C.D. (2000). *The workplace risk assessment* (WRA-20). Toronto, Ontario, Canada: Workplace.calm.

Bloom, H., Eisen, R.S., & Webster, C.D. (2001). *The employee risk assessment* (ERA-20). Toronto, Ontario, Canada: Workplace.calm.

Bloom, H., Webster, C., Hucker, S., & De Freitas, K. (2005). The Canadian contribution to violence risk assessment: History and implications for current psychiatric practice. *Canadian Journal of Psychiatry, 50,* 3-11.

*Blum, F. M. (2004). *Psychopathy, psychosis, drug abuse, and reoffense among conditionally released offenders.* Unpublished doctoral dissertation, University of Southern California, Los Angeles, CA.

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Sexual Violence Risk-20.* Burnaby, British Columbia, Canada: Simon Fraser University and British Columbia Forensic Psychiatric Services Commission.

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

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

Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines, and training. *American Psychologist, 51*, 945-956.

Borum, R., Bartel, P., & Forth, A. (2003). *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY)* (Version 1.1). Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute.

*Broner, N., Mayrl, D. W., & Landsberg, G. (2005). Outcomes of mandated and nonmandated New York City jail diversion for offenders with alcohol, drug, and mental disorders. *Prison Journal, 85*, 18-49.

*Brown, L. K. (2004). *Assessing risk for elopement and breaches of conditional release in insanity acquittees*. Unpublished doctoral dissertation, Simon Fraser University, Burnaby, British Columbia, Canada.

Buchanan, A. (1997). Assessing risk: Limits to the measurement of the target behaviours. *International Review of Psychiatry, 9,* 195-200.

Buchanan, A. (1999). Risk and dangerousness. *Psychological Medicine, 29*, 465-473.

Buchanan, A. (2001). [Review of the book *HCR-20: Assessing risk for violence, Version 2*]. *Criminal Behaviour and Mental Health, 11,* S77-S78.

Buchanan, A. (2008). Risk of violence by psychiatric patients: Beyond the "actuarial versus clinical" assessment debate. *Psychiatric Services, 59,* 184-190.

Buchanan, A., & David, A. (1994). Compliance and the reduction of dangerousness. *Journal of Mental Health, 3,* 427-429.

Buchanan, A., & Leese, M. (2001). Detention of people with dangerous severe personality disorders: A systemic review. *Lancet, 358,* 1955-1959.

Burgess, E. W. (1928). Factors determining success or failure on parole. In A. A. Bruce (Ed.), *The workings of the indeterminate sentence law and the parole system in Illinois* (pp. 205-249). Springfield, IL: Illinois Committee on Indeterminate-Sentence Law and Parole.

Campbell, J. (1995). Prediction of homicide of and by battered women. In J. D. Campbell (Ed.), *Assessing dangerousness* (pp. 96-113). Thousand Oaks, CA: Sage.

Campbell, J. C. (1986). Nursing assessment for risk of homicide with battered women. *Advances in Nursing Science, 8*, 36-51.

Campbell, M. A., French, S., & Gendreau, P. (2007). *Assessing the utility of risk assessment tools and personality measures in the prediction of violent recidivism for adult offenders* (User Report 2007-04). Ottawa, Ontario, Canada: Public Safety Canada.

Canadian Charter of Rights and Freedoms, Part 1 of the Constitution Act, 1982, being Schedule B to the Canada Act, 1982, (U.K.), 1982, c. 11U.S.C. (1982).

Canadian Foundation for Children, Youth and the Law v. Canada (Attorney General), [2004] 1 S.C.R. 76.

Canadian Psychological Association. (2000). *Canadian code of ethics for psychologists* (3rd ed.). Ottawa, Ontario, Canada: Author.

Canton, van der Veer, van Panhuis, Verheul, van den Brink (2004). De voorspellende waarde van risicotaxatie bij de rapportage pro Justitia. Onderzoek naar de HKT-30 en de klinische inschatting. *Tijdschrift voor psychiatrie, 46*, 525- 535.

*Catchpole, R. E. H., & Gretton, H. M. (2003). The predictive validity of risk assessment with violent young offenders: A 1-year examination of criminal outcome. *Criminal Justice and Behavior, 30*, 688-708.

Cauffman, E., & Steinberg, L. (2000a). (Im)maturity of judgment in adolescence: Why adolescents may be less culpable than adults. *Behavioral Sciences and the Law, 18,* 1-21.

Cauffman, E., & Steinberg, L. (2000b). Researching adolescents' judgment and culpability. In T. Grisso & R. G. Schwartz (Eds.), *Youth on trial: A developmental perspective in juvenile justice* (pp. 325-343). Chicago: University of Chicago Press.

Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the Health Professions, 25*, 12-37.

Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology, 70,* 6-20.

Claix, A., Pham, T., & Willocq, L. (2002, March). *Evaluation of the HCR-20 (historical-clinical-risk management) in a Belgian forensic population*. Poster presented at the annual conference of the international association of forensic mental health services, Munich, Germany.

Cocozza, J. J., & Steadman, H. J. (1976). The failure of psychiatric predictions of dangerousness: Clear and convincing evidence. *Rutgers Law Review, 29*, 1084-1101.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Lawrence Erlbaum Associates.

Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods, 8*, 243-253.

Cole, T. B., & Glass, R. M. (2005). Mental illness and violent death: Major issues for public health. *JAMA: Journal of the American Medical Association, 294*, 623-624.

Cooke, D. J., & Michie, C. (1998). Predicting recidivism in a Scottish prison sample. *Psychology, Crime & Law, 4*, 169-211.

Cooke, D. J., & Michie, C. (1999). Psychopathy across cultures: North America and Scotland compared. *Journal of Abnormal Psychology, 108*, 58-68.

Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological assessment, 13*, 171-188.

*Cooke, D. J., & Michie, C., & Ryan, J. (2001). Evaluating risk for violence: A preliminary study of the HCR-20, PCL-R and VRAG in a Scottish prison sample. *Scottish Prison Service occasional papers: no 5/200.*

Cooke, D. J., Michie, C., Hart, S. D., & Clark, D. (2005). Assessing psychopathy in the UK: Concerns about cross-cultural generalisability. *British Journal of Psychiatry, 186*, 335-341.

Coontz, P. D., Lidz, C. W., & Mulvey, E. P. (1994). Gender and the assessment of dangerousness in the psychiatric emergency room. *International Journal of Law & Psychiatry, 17*, 369-376.

*Cooper, S., Eslea, M., & Ireland, C. (July, 2008). *Psychopathy and institutional behaviour in adolescent offenders*. Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

Copas, J., & Marshall, P. (1998). The offender group reconviction scale: A statistical reconviction score for use by probation officers. *Applied Statistics, 47*, 159-171.

*Côté, G., Hodgins, S., & Daigle, M. (2001). *Violent behaviour, PCL-R and HCR-20 among involuntary inpatients, forensic patients and severely mentally disordered inmates.* Paper presented at the First Annual Meeting of the International Association of Forensic Mental Health Services, Vancouver, British Columbia, Canada.

Cottle, C. C., Lee, R. J., & Heilbrun, K. (2001). The prediction of criminal recidivism in juveniles: A meta-analysis. *Criminal Justice and Behavior, 28*, 367-394.

*Craig, L. A., Beech, A., & Browne, K. D. (2006). Cross-validation of the risk matrix 2000 sexual and violent scales. *Journal of Interpersonal Violence, 21*, 612-633.

*Craig, L. A., Browne, K. D., Beech, A., & Stringer, I. (2006). Differences in personality and risk characteristics in sex, violent and general offenders. *Criminal Behaviour & Mental Health, 16*, 183-194.

Criminal Code of Canada, Part XXIV (1977).

Crocker, A. G., Garcia, A., Israel, M., Hindle, Y., Gagnon, D., & Venegas, C. (n.d.). *Implementing and using a systematic risk assessment scheme to increase patient safety on a risk management unit for individuals with severe mental illness: A demonstration project.* Montreal, Québec, Canada: Douglas Mental Health University Institute.

*Dahle, K. (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.

Dawes, R. M. (2005). The ethical implications of Paul Meehl's work in comparing clinical versus actuarial prediction methods. *Journal of Clinical Psychology, 61*, 1245-1255.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81*, 95-106.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668-1674.

de Groot, A. D. (1961). Via clinical to statistical prediction. *Acta Psychologica, 18*, 274-284.

de Jonge, E., Nijman, H., Lammers, S., & Lucker, T. (2008). *Detecting changes in violence risk factors over time in a sample of forensic psychiatric patients in the Netherlands*. Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

*de Vogel, V., & de Ruiter, C. (2005). The HCR-20 in personality disordered female offenders: A comparison with a matched sample of males. *Clinical Psychology & Psychotherapy, 12*, 226-240.

*de Vogel, V., & de Ruiter, C. (2006). Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime & Law, 12*, 321-336.

*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.

*de Vogel, V., de Ruiter, C., van Beek, D., & Mead, G. (2004). Predictive validity of the SVR-20 and Static-99 in a Dutch sample of treated sex offenders. *Law and Human Behavior, 28*, 235-251.

de Vries Robbé, M., de Vogel, V., & de Spa, E. (2008). *The SAPROF: An instrument for the Structured Assessment of PROtective Factors for violence risk*. Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

*Dempster, R. J. (1998). *Prediction of sexually violent recidivism: A comparison of risk assessment instruments*. Unpublished master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada.

Dempster, R. J., & Hart, S. D. (2002). The relative utility of fixed and variable risk factors in discriminating sexual recidivists and nonrecidivists. *Sexual Abuse: A Journal of Research and Treatment, 14*, 121.

*Dernevik, M. (1998). Preliminary findings on reliability and validity of the Historical-Clinical-Risk Assessment in a forensic psychiatric setting. *Psychology, Crime, and Law, 4*, 127-137.

*Dernevik, M. (2004). *Professional risk assessment in clinical practice: The role of relevant clinical data gathering and context for the accuracy of structured professional risk assessment.* Stockholm, Sweden: Centre for Violence Prevention, Karolinska Institutet.

*Dernevik, M., Grann, M., & Johansson, 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.

Desmarais, S. L., Nicholls, T. L., & Brink, J. (2007). *User satisfaction and predictive validity of file-based Short-term Assessments of Risks: Comparisons with treatment team assessments.* Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Montréal, Québec, Canada.

*Dolan, M., & Fullam, R. (2007). The validity of the Violence Risk Scale 2nd edition (VRS-2) in a British Forensic Inpatient Sample. *The Journal of Forensic Psychology and Psychiatry, 18*, 381-193.

*Dolan, M., & Khawaja, A. (2004). The HCR-20 and post-discharge outcome in male patients discharged from medium security in the UK. *Aggressive Behavior, 30*, 469-483.

*Dolan, M. C., & Rennie, C. E. (2008). The Structured Assessment of Violence Risk in Youth as a predictor of recidivism in a United Kingdom cohort of adolescent offenders with conduct disorder. *Psychological Assessment, 20*, 35-46.

Douglas, K. S. (2001). *Making structured clinical decisions about violence risk: Reliability and validity of the HCR-20 violence risk assessment scheme.* Unpublished doctoral dissertation, Simon Fraser University, Burnaby, British Columbia, Canada.

Douglas, K. S., Cox, D. N., & Webster, C. D. (1999). Violence risk assessment: Science and practice. *Legal and Criminological Psychology, 4*, 149-184.

Douglas, K. S., Guy, L. S., & Weir, J. (2007). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography.* Retrieved December 5, 2008, from http://kdouglas.wordpress.com/hcr-20/hcr-20

Douglas, K. S., Hart, S. D., Webster, C. D., Belfrage, H., & Eaves, D. (in preparation). *HCR-20: Assessing risk for violence*, Version 3. Burnaby, British Columbia, Canada: Simon Fraser University.

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

Douglas, K. S., & Lavoie, J. (2006, March). *A model-based clustering approach to violence risk assessment using the HCR-20.* Paper presented at the meeting of the American Psychology-Law Society, St. Petersburg, Florida.

Douglas, K. S., Macfarlane, E., & Webster, C. D. (1996). Predicting dangerousness in the contemporary Canadian mental health and criminal justice systems. *Canada's Mental Health, 43*, 4-11.

*Douglas, K. S., & Ogloff, J. R. P. (2003a). Multiple facets of risk for violence: The impact of judgmental specificity on structured decisions about violence risk. *International Journal of Forensic Mental Health, 2*, 19-34.

Douglas, K. S., & Ogloff, J. R. P. (2003b). The impact of confidence on the accuracy of structured professional and actuarial violence risk judgments in a sample of forensic psychiatric patients. *Law and Human Behavior, 27*, 573-587.

Douglas, K. S., & Ogloff, J. R. (2003c). Violence by psychiatric patients: The impact of archival measurement source on violence base rates and risk assessment accuracy. *Canadian Journal of Psychiatry, 48*, 734-740.

*Douglas, K. S., Ogloff, J. R. P., & Hart, S. D. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services, 54*, 1372-1379.

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

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

Douglas, K. S., Strand, S., Belfrage, H., Fransson, G., & Levander, S. (2005). Reliability and validity evaluation of the psychopathy checklist: Screening version (PCL:SV) in Swedish correctional and forensic psychiatric samples. *Assessment, 12*, 145-161.

Douglas, K. S., & Webster, C. D. (1999). Predicting violence in mentally and personality disordered individuals. In R. Roesch, S. D. Hart, & J. R. P. Ogloff (Eds.), *Psychology and law: The state of the discipline* (pp. 175-239). New York: Plenum.

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

Douglas, K. S., Webster, C. D., Hart, S. D., Eaves, D., & Ogloff, J. R. P. (Eds.) (2001). *HCR-20: Violence risk management companion guide*. Burnaby, British Columbia, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University, and Department of Mental Health Law & Policy, University of South Florida.

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

*Dowsett, J. (2005). Measurement of risk by a community forensic mental health team. *Psychiatric Bulletin, 29*, 9-12.

Doyle, M., & Dolan, M. (2002). Violence risk assessment: Combining actuarial and clinical information to structure clinical judgements for the formulation and management of risk. *Journal of Psychiatric and Mental Health Nursing, 9*, 649-657.

*Doyle, M., & Dolan, M. (2006). Predicting community violence from patients discharged from community mental health services. *British Journal of Psychiatry, 189*, 520-526.

*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*, 141-152.

*Dunbar, E. (2003). Symbolic, relational, and ideological signifiers of bias-motivated offenders: Toward a strategy of assessment. *American Journal of Orthopsychiatry, 73*, 203.

*Dunbar, E., Quinones, J., & Crevecoeur, D. A. (2005). Assessment of hate crime offenders: The role of bias intent in examining violence risk. *Journal of Forensic Psychology Practice, 5*, 1-19.

Durlak, J. A. (1998). Common risk and protective factors in successful prevention programs. *American Journal of Orthopsychiatry, 68*, 512-520.

Dvoskin, J. A., & Heilbrun, K. (2001). Risk assessment and release decision-making: Toward resolving the great debate. *Journal of the American Academy of Psychiatry and the Law, 29*, 6-10.

Edens, J. F., & Campbell, J. S. (2007). Identifying youths at risk for institutional misconduct: A meta-analytic investigation of the psychopathy checklist measures. *Psychological Services, 4*, 13-27.

Edens, J. F., Campbell, J. S., & Weir, J. M. (2007). Youth psychopathy and criminal recidivism: A meta-analysis of the psychopathy checklist measures. *Law and Human Behavior, 31*, 53-75.

Edens, J. F., Skeem, J. L., & Douglas, K. S. (2006). Incremental validity analyses of the violence risk appraisal guide and the psychopathy checklist: Screening version in a civil psychiatric sample. *Assessment, 13*, 368-374.

Elbogen, E. B. (2002). The process of violence risk assessment: A review of descriptive research. *Aggression and Violent Behavior, 7*, 591-604.

Elbogen, E. B., Huss, M. T., Tomkins, A. J., & Scalora, M. J. (2005). Clinical decision making about psychopathy and violence risk assessment in public sector mental health settings. *Psychological Services, 2*, 133-141.

Elbogen, E. B., Williams, A. L., Kim, D., Tomkins, A. J., & Scalora, M. J. (2001). Gender and perceptions of dangerousness in civil psychiatric patients. *Legal and Criminological Psychology, 6*, 215-228.

*Enebrink, P., Långström, N., & Gumpert, C. H. (2006). Predicting aggressive and disruptive behavior in referred 6- to 12-year-old boys: Prospective validation of the EARL-20B risk/needs checklist. *Assessment, 13*, 356-367.

Enebrink, P., Långström, N., Gumpert, C. H., & Hulten, A. (2003). *Evaluating risk for antisocial behavior: Properties of the Early Assessment Risk List for Boys (EARL-20B)*. Huddinge, Sweden: Karolinska Institutet.

Enebrink, P., Långström, N., Hulten, A. & Gumpert, C.H. (2006). Swedish validation of the Early Assessment Risk List for Boys (EARL-20B), a decision aid for use with children presenting with conduct-disordered behaviour. *Nordic Journal of Psychiatry, 60*, 438-446.

Enebrink, P., Langström, N., Neij., J., Grann, M., & Gumpert, C. H. (2006). *Brief report: Interrater reliability of the Early Assessment Risk List: EARL-20B: A new guide for clinical evaluation of conduct-disordered boys.* Huddinge, Sweden: Karolinska Institutet.

Ennis, B. J., & Litwack, T. R. (1974). Psychiatry and the presumption of expertise: Flipping coins in the courtroom. *California Law Review, 62*, 693-752.

Epperson, D. L., Kaul, J. D., & Huot, S. J. (1995). *Predicting risk of recidivism for incarcerated sex offenders: Updated development on the Sex Offender Screening Tool (SOST).* Paper presented at the 14th annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, New Orleans, LA.

Epperson, D. L., Ralston, C. A., Fowers, D., & DeWitt, J. (2005). *Development of a sexual offense recidivism risk assessment tool–II (JSORRAT-II)*. Unpublished manuscript, University of Iowa, Ames, IA.

Ferguson, G. E., Eidelson, R. J., & Witt, P. H. (1998). New jersey's sex offender risk assessment scale: Preliminary validity data. *Journal of Psychiatry & Law, 26*, 327-351.

Field, A. P. (2003). The problems in using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics, 2*, 105-124.

*Fitch, D.C. (2002). *Analysis of common risk factors for violent behavior in Native American adolescents referred for residential treatment.* Unpublished doctoral dissertation, Texas Southern University, Houston, TX.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Flores, A. W., Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2006). Predicting outcome with the level of service inventory-revised: The importance of implementation integrity. *Journal of Criminal Justice, 34*, 523-529.

*Folino, J., Almirón, M., & Ricci, M. A. (2007). [Violent recidivism risk factor in filicidal women]. *Vertex* (Buenos Aires, Argentina), *18*, 258-267.

Folino, J. O., Marengo, C. M., Marchiano, S. E., & Ascazibar, M. (2004). The risk assessment program and the court of penal execution in the province of Buenos Aires, Argentina. *International Journal of Offender Therapy and Comparative Criminology, 48*, 49-58.

*Freese, R., Hiscoke, U., & Hodgins, S. (2002, May). *The treatment of mentally ill patients with a history of criminality or violence: What works and what doesn't work?* Paper presentation at 11th congress of the association of European psychiatrists, Stockholm, Sweden.

*Fujii, D., Lichton, A., & Tokioka, A. *Structured professional judgment versus actuarial data in violence risk prediction using the Historical Clinical Risk Management-20.* Manuscript submitted for publication.

*Fujii, D. E. M., 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.

Gardner, W., Lidz, C. W., Mulvey, E. P., & Shaw, E. C. (1996a). Clinical versus actuarial predictions of violence in patients with mental illness. *Journal of Consulting and Clinical Psychology, 64*, 602-609.

*Garman, G. (2004). *Violence, self harm and absconding in a secure psychiatric service*. Unpublished doctoral dissertation, University of Reading, Berkshire, United Kingdom.

Gelles, R., & Tolman, R. (1998). *The Kingston Screening Instrument for Domestic Violence (KSID)*. Unpublished risk instrument, University of Rhode Island, Providence, RI.

Gendreau, P., & Andrews, D. A. (1990). Tertiary prevention: What the meta-analyses of the offender treatment literature tell us about 'what works.' *Canadian Journal of Criminology, 32*, 173-184.

Gendreau, P., & Goggin, C. (1996). Principles of effective assessment for community corrections. *Federal Probation, 60*, 64-70.

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

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works. *Criminology, 34*, 575-607.

*Gibas, A. L., Kropp, P. R., & Hart, S. D. (2008). *Validity of the SARA in a Canadian sample of incarcerated adult males.* Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

Glaser, D. (1954). A reconsideration of some parole prediction factors. *American Sociological Review, 19*, 335-341.

Glaser, D. (1955a). Testing correctional decisions. *Journal of Criminal Law, Criminology & Police Science, 45*, 679-684.

Glaser, D. (1955b). The efficacy of alternative approaches to parole prediction. *American Sociological Review, 20*, 283-287.

Glaser, D. (1962). Prediction tables as accounting devices for judges and parole boards. *Crime & Delinquency, 8*, 239-258.

Glass, B. (1976a). The critical state of the critical review article. *Quarterly Review of Biology, 50th Anniversary Special Issue (1926-76)*, 415-418.

Glass, G. V (1976b). Primary, secondary and meta-analysis of research. *Educational Researcher, 10*, 3-8.

Gottfredson, S. D., & Gottfredson, D. M. (1988). Violence prediction methods: Statistical and clinical strategies. *Violence and Victims, 3*, 303-324.

Gottfredson, S. D., & Moriarty, L. J. (2006a). Clinical versus actuarial judgments in criminal justice decisions: Should one replace the other? *Federal Probation, 70*, 15-18.

Gottfredson, S. D., & Moriarty, L. J. (2006b). Statistical risk assessment: Old problems and new applications. *Crime & Delinquency, 52*, 178-200.

*Grann, M., Belfrage, H., & Tengstrom, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behavior, 27*, 97-114.

*Grann, M., Sturidsson, K., Haggard-Grann, U., Hiscoke, U. L., Alm, P., & Dernevik, M., et al. (2005). Methodological development: Structured outcome assessment and community risk monitoring (SORM). *International Journal of Law and Psychiatry, 28*, 442-456.

*Grann, M., & Wedin, I. (2002). Risk factors for recidivism among spousal assault and spousal homicide offenders. Psychology, *Crime and Law, 8*, 5-23.

*Gray, N. S., Fitzgerald, S., Taylor, J., MacCulloch, M. J., & Snowden, R. J. (2007). Predicting future reconviction in offenders with intellectual disabilities: The predictive efficacy of VRAG, PCL-SV and the HCR-20. *Psychological Assessment, 19,* 474-479.

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

*Gray, N. S., Snowden, R. J., MacCulloch, S., Phillips, H., Taylor, J., & MacCulloch, M. 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.

*Gray, N. S., Taylor, J., & Snowden, R. J. (2008) Predicting violent reconvictions using the HCR-20. British Journal of Psychiatry: *The Journal of Mental Science, 192,* 384-387.

Grebstein, L. C. (1963). Relative accuracy of actuarial prediction, experienced clinicians, and graduate students in a clinical judgment task. *Journal of Consulting Psychology, 27,* 127-132.

*Gretton, H.M., & Abramowitz, C.S. (2002, March). *The Structured Assessment of Violence Risk in Youth (SAVRY): Contribution of items and scales to clinical risk judgment and criminal outcomes.* Poster presented at the American Psychology and Law Society, Biennial Conference, Austin, Texas.

*Grevatt, M., Thomas-Peter, B., & Hughes, G. (2004). Violence, mental disorder and risk assessment: Can structured clinical assessments predict the short-term risk of inpatient violence? *Journal of Forensic Psychiatry & Psychology, 15,* 278-292.

Grisso, T. (1998). *Forensic evaluation of juveniles.* Sarasota, FL: Professional Resource Press.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2,* 293-323.

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.

Grubin, D. (1997). Predictors of risk in serious sex offenders. *British Journal of Psychiatry, 170,* 17-21.

Guy, L. S. (July, 2008). Performance of structured professional judgment tools for assessing risk of violence to others: A meta-analytic review. In J. Dvoskin (Chair), *Clarifying the principles of the structured professional judgment approach to violence risk assessment: Current status and future directions.* Symposium presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

Guy, L. S., & Douglas, K. S. (June, 2006). *HCR-20 violence risk assessment scheme: Evaluating item bias with item response theory.* Paper presented at the 6th annual conference of the International Association of Forensic Mental Health Services, Amsterdam, The Netherlands.

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.

Guy, L. S., & Wilson, C. M. (2007). *Empirical support for the HCR-20: A critical analysis of the violence literature*. Burnaby, British Columbia, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University. Retreived December 5, 2008, from http://kdouglas.files.wordpress.com/2007/09/hcr-20-report-2007.pdf

Haas, S. M., Hamilton, C. A., & Hanley, D. (2006). Implementation of the West Virginia Offender Reentry Initiative: An Examination of Staff Attitudes and the Application of the LSI-R. Charleston, WV: Mountain State Criminal Justice Research Services. Retreived December 5, 2008, from http://www.ojp.usdoj.gov/BJA/evaluation/psi_rp/wv-implementation.pdf

Hall, G. C. N. (1988). Criminal behavior as a function of clinical and actuarial variables in a sexual offender population. *Journal of Consulting and Clinical Psychology, 56*, 773-775.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29-36.

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

Hanson, R. K. (1997). *The development of a brief actuarial scale for sexual offense recidivism* (User report 1997-04). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

Hanson, R. K. (1998). What do we know about sex offender risk assessment? *Psychology, Public Policy, and Law, 4*, 50-72.

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

Hanson, R. K. (2006). Does Static-99 predict recidivism among older sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 18*, 343–356.

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

Hanson, R. K., Helmus, L., & Bourgon, G. (2007-draft version). *The Validity of Risk Assessments for Intimate Partner Violence: A Meta-Analysis*. Report submitted to Public Safety Canada, Ottawa, Ontario, Canada.

Hanson, R. K., & Morton-Bourgon, K. (2004). *Predictors of sexual recidivism: An updated meta-analysis*. Ottawa, Ontario, Canada: Public Safety and Emergency Preparedness Canada.

Hanson, R. K., & Morton-Bourgon, K. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*, 1154-1163.

Hanson, R. K., & Morton-Bourgon, K. E. (2007). *The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis*. Ottawa, Ontario, Canada: Public Safety and Emergency Preparedness Canada.

Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders*. Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static-2002*. (User report 2003-01). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

Hare, R. D. (2003). *The Revised Psychopathy Checklist*. Toronto, Ontario, Canada: Multi-Health Systems.

Harris, G. T., & Rice, M.E. (2007). Characterizing the value of actuarial violence risk assessments. *Criminal Justice and Behavior, 34*, 1638-1658.

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

Harris, G. T., Rice, M.E., & Quinsey, V.L. (2008). Shall evidence-based risk assessment be abandoned? *British Journal of Psychiatry, 192*, 154.

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

Hart, S. (2001). Assessing and managing violence risk. In K. Douglas, C. Webster, S. Hart, D. Eaves, & J. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 13-25). Burnaby, British Columbia, Canada: Mental Health, Law, & Policy Institute, Simon Fraser University, and Department of Mental Health Law and Policy, Florida Mental Health Institute, University of South Florida.

Hart, S. D. (2008, July). *The Structured Professional Judgment approach to violence risk assessment: Core principles of SPJ*. Paper presented at the annual meeting of the International Association of Forensic Mental Health Services, Vienna, Austria.

Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *Hare Psychopathy Checklist: Screening version*. Toronto, Ontario, Canada: Multi-Health Systems.

*Hart, S. D., Jackson, K., Healey, J., & Watt, K. A. (July, 2008). *Validation of the Risk for Sexual Violence Protocol in adult sexual offenders*. Symposium presented at the annual meeting of the International Association of Forensic Mental Health Services, Vienna, Austria.

Hart, S. D., Kropp, P. R., Laws, D. R., Klaver, J., Logan, C., & Watt, K. A. (2003). *The risk for sexual violence protocol (RSVP): Structured professional guidelines for assessing risk of sexual violence*. Vancouver, British Columbia, Canada: The Mental Health, Law, and Policy Institute, Simon Fraser University, Pacific Psychological Assessment Corporation, British Columbia Institute Against Family Violence.

*Hartvig, P., Alfarnes, S. A., Skjonberg, M., Moger, T. A., & Ostberg, B. (2006). Brief checklists for assessing violence risk among patients discharged from acute psychiatric facilities: A preliminary study. *Nordic Journal of Psychiatry, 60*, 243-248.

*Heckert, D. A., & Gondolf, E. W. (2004). Battered women's perceptions of risk versus risk factors and instruments in predicting repeat reassault. *Journal of Interpersonal Violence, 19*, 778-800.

Hedges, L. V. (1992). Meta-analysis. *Journal of Education Statistics, 17*, 279-296.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486-504.

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., Douglas, K. S., & Yasuhara, K. (in press). Violence Risk Assessment: Core Controversies. In J. Skeem, K. Douglas, & S. Lilienfeld (Eds.), *Psychological Science in the Courtroom: Controversies and Consensus*. New York: Guilford.

Heilbrun, K., Lee, R., & Cottle, C. C. (2005). Risk factors and intervention outcomes: Meta-analyses of juvenile offending. In K. Heilbrun, N. E. S. Goldstein & R. E. Redding (Eds.), *Juvenile delinquency: Prevention, assessment, and intervention.* (pp. 111-133). Oxford University Press.

Heilbrun, K., Nezu, C., Keeney, M., Chung, S., & Wasserman, A. (1998). Sexual offending: Linking assessment, intervention, and decision making. *Psychology, Public Policy, & Law, 4*, 138-174.

Heilbrun, K., O'Neill, M. L., Strohman, L. K., Bowman, Q., & Philipson, J. (2000). Expert approaches to communicating violence risk. *Law and Human Behavior, 24*, 137-148.

Hiday, V. A., Swartz, M. S., Swanson, J. W., Borum, R., & Wagner, H. R. (1998). Male-female differences in the setting and construction of violence among people with severe mental illness. *Social Psychiatry and Psychiatric Epidemiology, 33*, s68-s74.

Hildebrand, M., de Ruiter, C., & de Vogel, V. (2004). Psychopathy and sexual deviance in treated rapists: Association with sexual and nonsexual recidivism. *Sexual Abuse: Journal of Research and Treatment, 16*, 1-24.

Hildebrand, M., de Ruiter, C., & Nijman, H. (2004). PCL-R psychopathy predicts disruptive behavior among offenders in a Dutch forensic psychiatric hospital. *Journal of Interpersonal Violence, 19*, 13-29.

*Hildebrand, M., Hesper, B.L., Spreen, M., & Nijman, H.L.I. (2005). *De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie: een onderzoek naar de betrouwbaarheid en predictieve validiteit van de HCR-20, HKT-30 en PCL-R.* [The importance of structured risk assessment and of the diagnosis of psychopathy: A study into the reliability and predictive validity of the HCR-20, HKT-30 and PCL-R]. Utrecht, The Netherlands: Expertisecentrum Forensische Psychiatrie.

Hildebrand, M., Schönberger, H.J.M., & Spreen, M. (2007). *Onttrekkingen en recidives tijdens verlof gedurende de tbs-behandeling nader bekeken.* [A closer look at absconding and recidivism during forensic psychiatric treatment]. Utrecht, the Netherlands: Expertisecentrum Forensische Psychiatrie.

Hildebrand, M., Spreen, M., Schönberger, H.J.M., Augustinus, F., & Hesper, B.L. (2006). *Onttrekkingen tijdens verlof, ontvluchtingen, en recidives tijdens de tbs-behandeling in de jaren 2000-2005.* [Absconding during leave, escapes and recidivism during forensic psychiatric treatment between 2000-2005]. Utrecht, The Netherlands: Expertisecentrum Forensische Psychiatrie.

*Hill, A., Habermann, N., Klusmann, D., Berner, W., & Briken, P. (2008). Criminal recidivism in sexual homicide perpetrators. *International Journal of Offender Therapy and Comparative Criminology, 52*, 5-20.

*Hilterman, E. (2007, June). *Use of SAVRY by clinicians & its relation with recidivism by juveniles in Catalonia, Spain.* Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Montréal, Quebec.

Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *The Counseling Psychologist, 34*, 400-409.

*Hilton, N. Z., Harris, G. T., Rice, M. E., Houghton, R. E., & Eke, A. W. (2008). An indepth actuarial assessment for wife assault recidivism: The Domestic Violence Risk Appraisal Guide. *Law and Human Behavior, 32*, 150-163.

*Hilton, N. Z., Harris, G. T., Rice, M. E., Lang, C., Cormier, C. A., & Lines, K. J. (2004). A brief actuarial assessment for the prediction of wife assault recidivism: The Ontario Domestic Assault Risk Assessment. *Psychological Assessment, 16*, 267-275.

Hirsch, A. V. (1972). Prediction of criminal conduct and preventive confinement of convicted persons. *Buffalo Law Review, 21*, 717-758.

Hoffman, P. (1983). Screening for risk: A revised salient factor score (SFS 81). *Journal of Criminal Justice, 11*, 539-547.

Hoge, R. D. & D. A. Andrews. (2002). *Youth Level of Service/Case Management Inventory: User's manual.* Toronto, Ontario, Canada: Multi-Health Systems.

Hogue, T. E., & Allen, C. (2006). *Structured Clinical Judgement: Risk: Multidisciplinary assessment of risk. A system for structuring clinical judgements of risk related to high security hospital needs.* Unpublished manuscript, Rampton Hospital, Nottinghamshire Healthcare NHS Trust, Nottinghamshire, United Kingdom.

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.

Holt, R. R. (1970). Yet another look at clinical and statistical prediction: Or, is clinical psychology worthwhile? *The American Psychologist, 25*, 337-349.

Holt, R.R. (1986). Clinical and statistical prediction: A retrospective and would-be integrative perspective. *Journal of Personality Assessment, 54*, 376-386.

*Howard, B. (2007). *Examining predictive validity of the Salient Factor Score and HCR-20 among behavior health court clientele: Comparing static and dynamic variables.* Unpublished doctoral dissertation, Pacific Graduate School of Psychology, Palo Alto, CA.

Howells, K., Watt, B., Hall, G., & Baldwin, S. (1997). Developing programmes for violent offenders. *Legal and Criminological Psychology, 2*, 117-128.

Hrynkiw-Augimeri, L. K. (1998). *Assessing risk for violence in boys: A preliminary risk assessment study using the Early Assessment Risk List for Boys: EARL-20B.* Unpublished master's thesis, University of Toronto, Toronto, Ontario, Canada.

Hrynkiw-Augimeri, L. K. (2005). *Aggressive and antisocial young children: Risk assessment and management utilizing the early assessment risk list for boys (EARL-20B).* Unpublished doctoral dissertation, University of Toronto, Toronto, Ontario, Canada.

Hubbard, D. J., & Pratt, T. C. (2002). A meta-analysis of the predictors of delinquency among girls. *Journal of Offender Rehabilitation, 34*, 1-13.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*, 275-292.

Kahn, M. W. (1960). Clinical and statistical prediction revisited. *Journal of Clinical Psychology, 16*, 115-118.

Kansas v. Hendricks, 521 US 346 (1997).

*Kennealy, P. J., Skeem, J., Nicholson, E., Kregg, C., & Hart, E. (February, 2008). Risk Status or Risk State? Overlap between Problematic Personality Traits and Risk State Scores. In D. Murrie (Chair), *Violence risk assessment.* Symposium presented at the annual conference of the American Psychology-Law Society, Jacksonville, FL.

Koegl, C. J., Webster, C. D., Michel, M., & Augimeri, L. K. (2000). Coding raw data: Toward understanding raw life. *Child & Youth Care Forum, 29*, 229-246.

*Kreeves, K. A., Kropp, R., & Cairns, K. (2008). *An independent validation study of the SARA.* Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

*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.

*Kropp, P, R. (2002). *The Risk for Sexual Violence Protocol (RSVP).* Paper presented at the 21st Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Montréal, Québec, Canada.

Kropp, P. R. (2003). *Validity of law enforcement officers' evaluations of risk.* Paper presented at the 3rd Annual Conference of the International Association of Forensic Mental Health Services, Miami, FL.

Kropp, P. R., & Gibas, A. (in press). The Spousal Assault Risk Assessment Guide (SARA). In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment.* Oxford, United Kingdom: Routledge/Taylor & Francis.

*Kropp, P. R., & Hart, S. D. (2000). The Spousal Assault Risk Assessment (SARA) Guide: Reliability and validity in adult male offenders. *Law and Human Behavior, 24*, 101-118.

Kropp, P. R., & Hart, S. D. (2004). The development of the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER): A tool for criminal justice professionals. Family Violence Initiative, Research and Statistics Division, Department of Justice Canada.

Kropp, P. R., Hart, S. D., & Belfrage, H. (2005). *The Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER): User manual*. Vancouver, British Columbia, Canada: ProActive ReSolutions Inc.

Kropp, P. R., Hart, S. D., & Lyon, D. R. (2008). *The Stalking Assessment and Management Guidelines (SAM): User manual*. Vancouver, British Columbia, Canada: ProActive ReSolutions Inc.

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

Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1999). *Manual for the Spousal Assault Risk Assessment Guide*, 3rd ed. Toronto, Ontario, Canada: Multi-Health Systems.

Last, J. M. (2001). *A dictionary of epidemiology*. Oxford, United Kingdom: Oxford University Press.

Lavoie, J. A. A., Guy, L. S., & Douglas, K. S. (in press). Violence risk assessment: Principles and models bridging prediction to management. In J. Ireland (Ed.), *The assessment, treatment and management of violent and sexual offenders*. London, United Kingdom: Willan.

Leistico, A. R., Salekin, R. T., DeCoster, J., & Rogers, R. (2008). A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. *Law and Human Behavior, 32*, 28–45.

Lessard v. Schmidt, 349 F. Supp. 1078 (E.D. Wis. 1972).

Levene, K. S., Augimeri, L. K., Pepler, D., Walsh, M., Webster, C. D., & Koegel, C. J. (2001). *Early Assessment Risk List for Girls: EARL-21G*, Version 1 (consultation ed.). Toronto, Ontario, Canada: Earlscourt Child and Family Centre.

Levene, K. S., Walsh, M. M., Augimeri, L. K., & Pepler, D. (2004). Linking identification and treatment of early risk factors for female delinquency. In M. Moretti, C. Odgers, & M. Jackson (Eds.), *Girls and aggression: Contributing factors and intervention principles: Perspectives in law and psychology* (Vol. 19; pp. 147-164). New York: Kluwer Academic/Plenum Publishers.

Lidz, C. W., Mulvey, E. P., & Gardner, W. (1993). The accuracy of predictions of violence to others. *Journal of the Medical Association, 269*, 1007-1111.

Lindsay, W. R., & Beail, N. (2004). Risk assessment: Actuarial prediction and clinical judgement of offending incidents and behaviour for intellectual disability services. *Journal of Applied Research in Intellectual Disabilities, 17*, 229-234.

Lindsay, W. R., Murphy, L., Smith, G., Murphy, D., Edwards, Z., Grieve, A., et al. (2004). The dynamic risk assessment and management system: An assessment of immediate risk of violence for individuals with offending and challenging behaviour. *Journal of Applied Research in Intellectual Disabilities, 17*, 267–274.

*Lindsay, W., Hogue, T., Taylor, J., Steptoe, L., Mooney, P., O'Brien, G., et al. (2008). Risk assessment in offenders with intellectual disability: A comparison across three levels of security. *International Journal of Offender Therapy and Comparative Criminology, 52*, 90-111.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications, Inc.

Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law, 7*, 409-443.

*Lodewijks, de Ruiter, & Doreleijers (in press). Gender differences in violent outcome and risk assessment in adolescent offenders after residential treatment. *International Journal of Forensic Mental Health.*

*Lodewijks, H. P. B., Doreleijers, Th. A. H., & Ruiter, C. de (2008). SAVRY risk assessment in relation to sentencing and subsequent recidivism in a Dutch sample of violent juvenile offenders. *Criminal Justice and Behavior, 35*, 696-709.

*Lodewijks, H. P. B., Doreleijers, Th. A. H., Ruiter, C. de, & Borum, R. (2008). Predictive validity of the Structured Assessment of Violence in Youth (SAVRY) during residential treatment. *International Journal of Law and Psychiatry, 31*, 263-271.

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., et al. (1999). The researcher's own therapy allegiances: A "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*, 95-106.

Lyon, D. R., Hart, S. D., & Webster, C. D. (2001). Violence and risk assessment. In R. A. Schuller, & J. R. P. Ogloff (Eds.), *Introduction to psychology and law: Canadian perspectives* (pp. 314-350). Toronto, Ontario, Canada: University of Toronto Press.

Lyons v. the Queen, 37 C.C.C. (3d) 1 at 47 (S.C.C.) (1987).

*MacPherson, G. J. D., & Kevan, I. (2004). Predictive validity of the HCR-20 violence risk assessment scheme within a maximum security special hospital. *Issues in Forensic Psychology, 5*, 62-80.

*McCoy, W. K. (2007). *Predicting treatment outcome and recidivism among juvenile sex offenders: The utility of the JSOAP-II and ERASOR in an outpatient treatment program.* Unpublished doctoral dissertation, Sam Houston State University, Huntsville, TX.

*McDermott, B. E., Edens, J. F., Quanbeck, C. D., Busse, D., & Scott, C. L. (2008). Examining the role of static and dynamic risk factors in the prediction of inpatient violence: Variable- and person-focused analyses. *Law and Human Behavior, 34*, 325-338.

*McEachran, A. K. (2001). *The predictive validity of the PCL:YV and the SAVRY in a population of adolescent offenders*. Unpublished master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada.

*McGowan, M. (2007). *The predictive validity of violence risk assessment within educational settings*. Unpublished doctoral dissertation, Northern Arizona University, Flagstaff, AZ.

McGrath, R. J. & Hoke, S. E. (2001). *Vermont Assessment of Sex Offender Risk Manual*, Research Edition. Waterbury, VT: Author.

*McKenzie, B., & Curr, H. (2005). Predicting violence in a medium secure setting: A study using the historical and clinical scales of the HCR-20. *British Journal of Forensic Practice, 7*, 22-28.

McNiel, D. E. (1998). Empirically based clinical evaluation and management of the potentially violent patient. In P. M. Kleespies (Ed.), *Emergencies in mental health practice: Evaluation and management.* (pp. 95-116). Guilford Press.

McNiel, D. E., & Binder, R. L. (1995). Correlates of accuracy in the assessment of psychiatric inpatients' risk of violence. *American Journal of Psychiatry, 152*, 901-906.

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

Meehl, P. E. (1954). Clinical versus statistical prediction: A theoretical analysis and review of the evidence. Minneapolis, MN: The University of Minneapolis Press.

Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (2007). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (3rd ed.). NewYork: Guilford.

Metz, C. (1978). Basic Principles of ROC Analysis. *Seminars in Nuclear Medicine, 8*, 283-298.

*Meyers, J. R., & Schmidt, F. (2008). Predictive validity of the Structured Assessment for Violence Risk in Youth (SAVRY) with juvenile offenders. *Criminal Justice and Behavior, 35*, 344-355.

Miller, M., & Morris, N. (1988). Predictions of dangerousness: An argument for limited use. *Violence and Victims, 3*, 263-283.

Mills, J. F. (2005). Advances in the assessment and prediction of interpersonal violence. *Journal of Interpersonal Violence, 20*, 236-241.

*Mills, J., Kroner, D., & Hemmati, T. (2007). The Validity of Violence Risk Estimates: An Issue of Item Performance. *Psychological Services, 4*, 1-12.

Moffitt, T. E., Caspi, A., Dickson, N., Silva, P., & Stanton, W. (1996). Childhood-onset versus adolescent-onset antisocial conduct problems in males: Natural history from ages 3 to 18 years. *Development and Psychopathology, 8*, 399-424.

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology, 13*, 355-375.

Monahan, J. (1981). Predicting violent behavior: An assessment of clinical techniques. Beverley Hills, CA: Sage.

Monahan, J. (1996). Violence prediction: The last 20 years and the next 20 years. *Criminal Justice and Behavior, 23*, 107-120.

Monahan, J. (2008). Structured risk assessment of violence. In R. I. Simon, & K. Tardiff (Eds.), *Textbook of Violence Assessment and Management* (pp. 17-33). Arlington, VA: American Psychiatric Publishing.

Monahan, J., & Steadman, H. J. (2001). Violence risk assessment: A quarter century of research. In L. E. Frost, & R. J. Bonnie (Eds.), *The evolution of mental health law*. (pp. 195-211). Washington, DC: American Psychological Association.

Monahan, J., Steadman, H. J., & Appelbaum, P. S. (2001). Developing a clinically useful actuarial tool for assessing violence risk. *British Journal of Psychiatry, 176*, 299-312.

Monahan, J., Steadman, H. J., Robbins, P. C., Appelbaum, P., Banks, S., & Grisso, T., et al. (2005a). An actuarial model of violence risk assessment for persons with mental disorders. *Psychiatric Services, 56*, 810-815.

Monahan, J., Steadman, H. J., Silver, E., Applebaum, P. S., Robbins, P. C., & Mulvey, E. P. et al. (2001b). *Rethinking risk assessment: The MacArthur study of mental disorder and violence.* New York, NY: Oxford University Press.

Monahan, J., Steadman, H., Appelbaum, P., Grisso, T., Mulvey, E., Roth, L., et al. (2005). *Classification of Violence Risk: Professional manual.* Lutz, FL: PAR.

Monahan, J., Steadman, H., Robbins, P. C., Appelbaum, P., Banks, S., Grisso, T., et al. (2005b). Prospective validation of the multiple iterative classification tree model of violence risk assessment. *Psychiatric Services, 56*, 810-815.

*Morrissey, C., Hogue, T., Mooney, P., Allen, C., Johnston, S., Hollin, C., Lindsay, W. R,. & Taylor, J. L. (2007). Predictive validity of the PCL-R in offenders with intellectual disability in a high secure hospital setting: Institutional aggression. *Journal of Forensic Psychiatry & Psychology, 18*, 1-15.

*Morton, K. (2003). Psychometric properties of four risk assessment measures with male adolescent sexual offenders. Unpublished master's thesis, Carleton University, Ottawa, Ontario, Canada.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783-792.

Mossman, D. (2000). Evaluating violence risk 'by the book': A review of HCR-20: Assessing risk for violence, Version 2 and The manual for the Sexual Violence Risk-20. *Behavioral Sciences and the Law, 18,* 781-789.

*Mowat-Leger, V. (2001). Risk factors for violence: A comparison of domestic batterers and other violent and non-violent offenders. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.

*Müller-Isberner, J. 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.

Müller-Isberner, R., & Jockel, D. (1997, September). *The implementation of the HCR-20 in a German hospital order institution*. Paper presented at the 7th European Conference on Psychology and Law, Solna, Sweden.

Mulvey, E. P. (2005). Risk assessment in juvenile justice policy and practice. In K. Heilbrun, N. E. Sevin Goldstein, & R. Redding (Eds.), *Juvenile delinquency: Prevention, assessment, and intervention* (pp. 209-231). New York, NY: Oxford University Press.

Mulvey, E. P., Shaw, E., & Lidz, C. W. (1994). Why use multiple sources in research on patient violence in the community? *Criminal Behaviour and Mental Health, 4*, 253-258.

Mulvey, E., & Lidz, C. (1995) Conditional prediction: A model for research on dangerousness to others in a new era. *International Journal of Law & Psychiatry, 18*, 129-143.

Mulvey, E., Lidz, C., Gardner, W., & Shaw, E. (1996). Clinical versus actuarial predictions of violence in patients with mental illnesses. *Journal of Consulting & Clinical Psychology, 64*, 602-609.

National Mental Health Risk Management Programme, Department of Health (2007) *Best Practice in Managing Risk: Principles and Evidence for Best Practice in the Assessment and Management of Risk to Self and Others in Mental Health Services.* London: Department of Health.

*Neves, A. C., & Gonçalves, R. A. (July, 2008). Criminal recidivism and violation of conditional release: A comparative study of the LSI-R and the HCR-20 with Portuguese Probationers/parolees. Poster presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

*Nicholls, T. L. (2001). *Violence risk assessments with female NCRMD acquittees: Validity of the HCR-20 and PCL-SV.* Unpublished doctoral dissertation, Simon Fraser University, Burnaby, British Columbia, Canada.

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

*Nicholls, T. L., Ogloff, J. R. P., & Douglas, K. S. (2004). Assessing risk for violence among female and male civil psychiatric patients: The HCR-20, PCL: SV, and McNiel & Binder's VSC. *Behavioral Sciences and the Law, 22*, 127-158.

Nikolova, N., Collins, M. J., Guy, L. S., Lavoie, J., Reeves, K., Wilson, C. M., et al. (March, 2006). *HCR-20 violence risk assessment scheme: A quantitative synthesis of its application, reliability and validity*. Poster presented at the annual conference of the American Psychology-Law Society, St. Petersburg, Florida.

Nuffield, J. (1982). Parole decision-making in Canada: Research towards decision guidelines. Ottawa, Ontario, Canada: 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, ON.

Odgers, C. L., Schmidt, M. G., & Reppucci, N. D., (2004). Reframing violence risk assessment for female juvenile offenders. In M. M. Moretti, C. L. Odgers, & M. Jackson (Eds.), *A Girls and aggression: Contributing factors and intervention principles*. (pp. 195-210). New York: Kluwer Academic/Plenum Publishers.

Odgers, C. L., Vincent, G. M., Corrado, R. R. (2002). A preliminary conceptual framework for the prevention and management of multi-problem youth. In R. Corrado, R. Roesch, S. D. Hart, & J. K. Gierowski (Eds.), *Multi-problem violent youth: A foundation for comparative research on needs, interventions and outcomes.* (pp. 302-329). Amsterdam, Netherlands: IOS Press.

Ogloff, J. R. P.& Daffern, M. (2006). The dynamic appraisal of situational aggression: An instrument to assess risk for imminent aggression in psychiatric inpatients. *Behavioral Sciences & the Law, 24*, 799-813.

Otto, R. K. (2000). Assessing and managing violence risk in outpatient settings. *Journal of Clinical Psychology, 56*, 1239-1262.

Packer, I. K. (2001). *Use of Violence Risk Assessment Instruments: A National Survey*. Paper proposal submitted to the annual conference of the American Psychology-Law Society, Austin, Texas.

Patterson G.R., Forgatch M.S., Voerger K.L., & Stoolmiller M. (1998). Variables that initiate and maintain an early-onset trajectory of offending. *Developmental Psychopathology, 10*, 531-547.

Penetanguishene Mental Health Centre v. Ontario (Attorney General), [2004] 1 S.C.R. 498.

*Penney, S. R., Lee, Z. L., Moretti, M. M., & Bartel, P. (2007). *The predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY)*. Poster presented at the annual conference of the Forensic Psychiatric Services Commission, Victoria, British Columbia, Canada.

*Pham, T. H., Chevrier, I., Nioche, A., Ducro, C., & Réveillère, C. (2005). Psychopathie, évaluation du risque, prise en charge [Psychopathy, risk assessment and rehabilitation]. *Annales Médico-Psychologiques, 163*, 878-881.

*Pham, T. H., Ducro, C., Marghem, B., & Réveillère, C. (2005). Évaluation du risque de récidive au sein d'une population de délinquants incarcérés ou internés en Belgique francophone [Prediction of recidivism among prison inmates and forensic patients in Belgium]. *Annales Médico-Psychologiques, 163*, 842-845.

*Philipse, M. (2002, March). *Post-dictive validity of the HCR-20 in a Dutch forensic psychiatric sample.* Paper presented at the Annual Conference of the International Association of Forensic Mental Health Services, Munich, Germany.

Philipse, M. W. G., Koeter, M. W. J., van den Brink, W., & van der Staak,Cees P.F. (2004). The structural coherence of clinically derived dynamic indicators of reoffending risk. *Criminal Behaviour & Mental Health, 14*, 263-279.

*Polvi, N. H. (1999). 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, Burnaby, British Columbia, Canada.

Prentky, R., & Righthand, S. (2003). *Juvenile sex offender assessment protocol-II (J-SOAP-II) manual*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Quinsey, V. L., & Maguire, A. (1986). Maximum security psychiatric patients: Actuarial and clinical predictions of dangerousness. *Journal of Interpersonal Violence, 1*,143-171.

Quinsey, V. L., Book, A., & Skilling, T. A. (2004). A follow-up of deinstitutionalised developmentally handicapped men with histories of antisocial behaviour. *Journal of Applied Research in Intellectual Disabilities, 17*, 243-254.

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

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

R. v. Johnson, 2003, 2 S.C.R. 357.

*Ramirez, M. P., Illescas, S. R., Garcia, M. M., Forero, C. G., & Pueyo, A. A. (2008). Prediccion de riesgo de reincidencia en agresores sexuales. *Psicothema, 20*, 205-210.

Re Moore and the Queen, 1984.

Reeves, K., Nikolova, N., Guy, L. S., Collins, M., Wilson, C., Lavoie, J. A. A., et al. (in preparation). HCR-20 violence risk assessment scheme: A quantitative synthesis of its application, reliability and validity.

Rice, M. E. & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737-748.

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

Rice, M. E., Harris, G. T., & Hilton, N. Z. (in press). The Violence Risk Appraisal Guide and Sex Offender Risk Appraisal Guide for Violence Risk Assessment and the Ontario Domestic Assault Risk Assessment and Domestic Violence Risk Appraisal Guide for Wife Assault Risk Assessment. In R. K. Otto, & K. S. Douglas (Eds.), *Handbook of violence risk assessment*. Oxford, United Kingdom: Routledge/Taylor & Francis.

Rice, M.E. (1997). Violent offender research and implications for the criminal justice system. *American Psychologist, 52*, 414-423.

Richards, J .M. (1963). Reconceptualization of the clinical and statistical prediction controversy in terms of components of accuracy of interpersonal perception scores. *Psychological Reports, 12*, 443-448.

Risk Management Authority Scotland. (2007). *Risk Assessment Tools Evaluation Directory* (Version 2). Paisley: Author.

Robbins, P. C., Monahan, J., & Silver, E. (2003). Mental disorder, violence, and gender. *Law and Human Behavior, 27*, 561-571.

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

Rosenfeld, B. (2004). Violence risk factors in stalking and obsessional harassment: A review and preliminary meta-analysis. *Criminal Justice and Behavior, 31*, 9-36.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638-641.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*, 59-82.

*Ross, D. J., Hart, S. D., & Webster, C. D. (1998). *Facts and fates: Prediction and management of hospital and community aggression using the HCR-20*. Port Coquitlam, British Columbia: Riverview Hospital.

Sarbin, T. (1943). A Contribution to the Study of Actuarial and Individual Methods of Prediction. *American Journal of Sociology, 48*, 593-602.

Sarbin, T. R. (1962). The present status of the clinical-statistical prediction problem. *Anthropology & Medicine, 10*, 315-323.

Schall v. Martin, 467 US 253 (1984)

*Schönberger, H. J. M., Hildebrand, M., Spreen, M., & Bloem, O. (2008). *De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie bij seksuele delinquenten: een onderzoek naar de betrouwbaarheid en predictieve validiteit van de SVR-20, Static-99, HKT-30 en PCL-R* [The importance of structured risk assessment and of the diagnosis of psychopathy in sex offenders: A study into the reliability and predictive validity of the SVR-20, Static-99, HKT-30 en PCL-R]. Utrecht, the Netherlands: Expertisecentrum Forensische Psychiatrie.

Schopp, R. F. (1996). Communicating risk assessments: Accuracy, efficacy, and responsibility. *American Psychologist, 51*, 939-944.

Schwalbe, C. S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and Human Behavior, 31*, 449-462.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261-281). New York: Russell Sage Foundation.

Shah, S. A. (1978). Dangerousness: A paradigm for exploring some issues in law and psychology. *American Psychologist, 33*, 224-238.

Sines, J. O. (1970). Actuarial versus clinical prediction in psychopathology. The British Journal of Psychiatry: *The Journal of Mental Science, 116*, 129-144.

*Sjösted, G., & Långström, N. (2002). Assessment of risk for criminal recidivism among rapists: A comparison of four different measures. *Psychology, Crime, & Law, 8*, 25-40.

Skeem, J. L., Monahan, J., & Mulvey, E. P. (2002). Psychopathy, treatment involvement, and subsequent violence among civil psychiatric patients. *Law and Human Behavior, 26*, 577-603.

Skeem, J. L., Mulvey, E. P., Odgers, C., Schubert, C., Stowman, S., & Gardner, W. et al. (2005). What do clinicians expect? comparing envisioned and reported violence for male and female patients. *Journal of Consulting and Clinical Psychology, 73*, 599-609.

Skeem, J., Schubert, C., Stowman, S., Beeson, S., Mulvey, E., & Gardner, W. et al. (2005). Gender and risk assessment accuracy: Underestimating women's violence potential. *Law and Human Behavior, 29*, 173-186.

*Skowron, C. (2004). *Differentiation and predictive factors in adolescent sexual offending.* Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.

Smith v. Jones, 1999, 1 S.C.R. 455.

Smith, J., & Lanyon, R. I. (1968). Prediction of juvenile probation violators. *Journal of Consulting and Clinical Psychology, 32*, 54-58.

*Stadtland, C., & Nedopil, N. (2005). [Psychiatric disorders and the prognosis for criminal recidivism]. *Der Nervenarzt, 76*, 1402-1411.

*Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U., & Nedopil, N. (2005). Risk assessment and prediction of violent and sexual recidivism in sex offenders: Long-term predictive validity of four risk assessment instruments. *Journal of Forensic Psychiatry & Psychology, 16*, 92-108.

Steadman, H. J. (2000). From dangerousness to risk assessment of community violence: Taking stock at the turn of the century. *Journal of the American Academy of Psychiatry and the Law, 28*, 265-271.

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

Steinberg, L., & Cauffman, E. (1996). Maturity of judgment in adolescence: Psychosocial factors in adolescent decision making. *Law and Human Behavior, 20*, 249-272.

Steinberg, L., & Cauffman, E. (1999). A developmental perspective on serious juvenile crime: When should juveniles be treated as adults? *Federal Probation, 63*, 52-57.

Strand, S., & Belfrage, H. (2001). Comparison of HCR-20 scores in violent mentally disordered men and women: Gender differences and similarities. *Psychology, Crime & Law, 7*, 71-79.

*Strand, S., Belfrage, H., Fransson, G., & Levander, 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, 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, Knoxville, TN.

Stricker, G. (1967). Actuarial, naive clinical, and sophisticated clinical prediction of pathology from figure drawings. *Journal of Consulting Psychology, 31*, 492-494.

Swets, J. A. (1986). Indices of discrimination and diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99*, 100-177.

Swets, J. A. (1996). Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers. Mahwah, NJ: Lawrence Erlbaum Associates.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1-26.

Tape, T. (n.d.). The area under the ROC curve. Web based manuscript. Available: http://gim.unmc.edu/dxtests/ROC3.htm.

Tarasoff v. Regents of University of California (1976) 17 Cal. 3d 425

Teasdale, B., Silver, E., & Monahan, J. (2006). Gender, Threat/Control-Override Delusions and Violence. *Law and Human Behavior, 30*, 649-658.

*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.

*Tengstrom, A., Hodgins, S., Muller-Isberner, R., Jockel, D., Freese, R., & Ozokyay, K., et al. (2006). Predicting violent and antisocial behavior in hospital using the HCR-20: The effect of diagnoses on predictive accuracy. *International Journal of Forensic Mental Health, 5*, 39-53.

Thorne, F. C. (1960). Clinical judgment: A clinician's viewpoint. *Journal of Clinical Psychology, 16*, 128-134.

Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. Sexual Abuse: *A Journal of Research & Treatment, 14*, 139-153.

Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., et al. (2003). Distinguishing and combining risks for sexual and violent recidivism. *Annals of the New York Academy of Sciences, 989*, 225-235.

Tilt R, Perry B, Martin C, Maguire M, Preston M (2000) *Report of the Review of Security at the High Security Hospitals*. London: Department of Health.

*Urheim, R., Jakobsen, D., & Rasmussen, K. (2003, August). *Dimensions of inpatient aggressive behavior in a security ward: What is being "predicted"?* Paper presented at the 5th Nordic Symposium on Forensic Psychiatry, Ystad, Sweden.

van den Broek, E., & de Vries Robbé, M. (2008). *The supplemental value of the SAPROF from a treatment perspective: A counterbalance to risk?* Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

van Marle, H. J. C. (2002). The Dutch Entrustment Act (TBS): its principals and innovations. *International Journal of Forensic Mental Health, 1*, 83-92.

*Viljoen, J. L., Scalora, M., Cuadra, L., Bader, S., Chavez, V., Ullman, D., et al. (2008). Assessing risk for violence in adolescents who have sexually offended: A comparison of the J-SOAP-II, J-SORRAT-II, and SAVRY. *Criminal Justice and Behavior, 35*, 5-23.

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

*Warren, J. I., South, S. C., Burnette, M. L., Rogers, A., Friend, R., Bale, R., & Van Patten, I. (2005). Understanding the risk factors for violence and criminality in women: The concurrent validity of the PCL-R and HCR-20. *International Journal of Law and Psychiatry, 28*, 269-289.

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

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Middleton, C. (2004). *Manual for the Short-Term Assessment of Risk and Treatability (START)*, Version 1.0 (Consultation Ed.). Hamilton, ON: St. Joseph's Healthcare Hamilton and Port Coquitlam, BC: Forensic Psychiatric Services Commission.

Webster, C. D., Müller-Isberner, R., & Fransson, G. (2002). Violence risk assessment: Using structured clinical guides professionally. *International Journal of Forensic Mental Health, 1*, 185-193.

Webster, C. D., Douglas, K., Eaves, D. & Hart, S. (1997). *HCR-20: Assessing risk for violence*, Version 2. Burnaby, British Columbia, Canada: Simon Fraser University.

Webster, C. D., Nicholls, T. L., Martin, M. L., Desmarais, S. L., & Brink, J. (2006). Short-Term Assessment of Risk and Treatability (START): The case for a new structured professional judgment scheme. *Behavioral Sciences & the Law, 24*, 747-766.

Webster, S. D., Mann, R. E., Carter, A. J., Long, J., Milner, R. J., O'Brien, et al. (2006). Inter-rater reliability of dynamic risk assessment with sexual offenders. *Psychology, Crime and Law, 12*, 439–452.

*Welsh, J., Schmidt, F., McKinnon, L., Chattha, H., & Meyers, J. (2008). A comparative study of adolescent risk assessment instruments: Predictive and incremental validity. *Assessment, 15*, 104-115.

Werkgroep Pilotstudy Risicotaxatie Forensische Psychiatrie. (2002). *Bevindingen van een landelijke pilotstudy naar de HKT-30*. Den Haag: Ministerie van Justitie.

*Williams, K, R., & Houghton, A. B. (2004). Assessing the risk of domestic violence reoffending: A validation study. *Law and Human Behavior, 28*, 437-455.

*Wintrup, A. (1996). *Assessing risk of violence in mentally disordered offenders with the HCR-20.* Unpublished master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada.

Witt, P. H. (2000). A practitioner's view of risk assessment: The HCR-20 and SVR-20. *Behavioral Sciences and the Law, 18*, 791-798.

Wong, S., & Gordon, A. (2000). *Manual for the Violence Risk Scale.* Saskatoon: SK: University of Saskatchewan.

Wootton, L., Buchanan, A., Leese, M., Tyrer, P., Burns, T., Creed, F., Fahy, T., & Walsh, E. (2008). Violence in psychosis: estimating the predictive validity of readily accessible clinical information in a community sample. *Schizophrenia Research, 101*, 176-184.

Worling, J. R., & Curwen, T. (2001). *Estimate of Risk Adolescent Sexual Offense Recidivism (ERASOR) Version 2.0*. Toronto, Ontario, Canada: SAFE-T Program, Thistletown Regional Centre, Ontario Ministry of Community and Social Services.

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

Yeaton, W. H., & Wortman, P. M. (1993). On the reliability of meta-analytic reviews: The role of intercoder agreement. *Evaluation Review, 17*, 292-309.

Young, D., Moline, K., Farrell, J., & Bierie, D. (2006). Best implementation practices: Disseminating new assessment technologies in a juvenile justice agency. *Crime & Delinquency, 52*, 135-158.

Yuile, A. L. (2008). A gender-sensitive approach to girlhood aggression: Risk assessment, intervention, and follow-up. Unpublished doctoral dissertation, York University, Toronto, Ontario, Canada.

APPENDICES

Appendix 1.1.

Scales and Items in the HCR-20

Scale	Item
Historical	
1	Previous Violence
2	Young Age at First Violent Incident
3	Relationship Instability
4	Employment Problems
5	Substance Use Problems
6	Major Mental Illness
7	Psychopathy
8	Early Maladjustment
9	Personality Disorder
10	Prior Supervision Failure
Clinical	
1	Lack of Insight
2	Negative Attitudes
3	Active Symptoms of Major Mental Illness
4	Impulsivity
5	Unresponsive to Treatment
Risk Management	
1	Plans Lack Feasibility
2	Exposure to Destabilizers
3	Lack of Personal Support
4	Noncompliance with Remediation Attempts
5	Stress

Adapted from Webster, Douglas, Eaves, and Hart (1997).

Appendix 1.2.

Items in the START

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

Adapted from Webster, Martin, Brink, Nicholls, and Middleton (2004).

Appendix 1.3.

Scales and Items in the SCJ: Risk

Scale	Item
Historical	
1	Previous Violence
2	Young Age at First Violent Incident
3	Relationship Instability
4	Employment Problems
5	Substance Use Problems
6	Major Mental Illness
7	Psychopathy
8	Early Maladjustment
9	Personality Disorder
10	Prior Supervision Failure
11	Child Protection
12	Sex Offending
13	Suicide Attempt
14	Self-Harm
15	Arson
16	Hostage Taking
17	Weapons
18	Concerted Indiscipline
19	High Public or Political Interest
20	Escape/Abscond History
Clinical	
1	Lack of Insight
2	Negative Attitudes
3	Active Symptoms of Major Mental Illness
4	Impulsivity
5	Unresponsive to Treatment
Suicide Attempt or Self-Harm Items	
1	Suicidal Ideation
2	Hopelessness
3	Frequency
4	Severity
5	Planning

Scale	Item
Vulnerability	
1	Mental State
2	Physical/Physiological Problems
3	Psychological Problems
4	Social Problems
5	Exploitation
Escape and Subversion	
1	Planning
2	Incentive
3	Interest in Security
4	Mental Disorder
5	Subversive Behaviour
Risk Management	
1	Plans Lack Feasibility
2	Exposure to Destabilizers
3	Lack of Personal Support
4	Noncompliance with Remediation Attempts
5	Stress

Adapted from Hogue and Allen (2006).

Appendix 1.4.

Scales and Items in the HKT-30

Scale	Item
Historical	
1	Judicial History
2	Violation of Conditions
3	Conduct Problems Before the Age of Twelve
4	Victim of Violence in Youth
5	History of Social or Mental Health Care
6	History of Employment
7	Use of Drugs
8	Psychotic Disorder
9	Personality Disorder
10	Psychopathy
11	Sexual Deviance
Clinical	
1	Lack of Insight
2	Psychotic Symptoms
3	Use of Drugs
4	Impulsivity
5	Empathy
6	Hostility
7	Social and Relational Skills
8	Ability to Look After Oneself
9	Acculturation Problems
10	Attitude Towards Treatment
11	Responsibility Towards Crime
12	Sexual Preoccupation
13	Coping Skills
Future	
1	Agreement on Conditions
2	Material Indicators
3	Daily Activities
4	Skills
5	Social Support and Network
6	Exposure to Destabilizers

Adapted from Werkgroep Risicotaxatie Forensische Psychiatrie (2002).

Appendix 1.5.

Scales and Items in the SVR-20

Scale	Item
Psychosocial Adjustment	
1	Sexual Deviation
2	Victim of Child Abuse
3	Psychopathy
4	Major Mental Illness
5	Substance Use Problems
6	Suicidal/Homocidal Ideation
7	Relationship Problems
8	Employment Problems
9	Past Nonsexual Violent Offences
10	Past Nonviolent Offences
11	Past Supervision Failure
Sexual Offences	
12	High Density Sex Offences
13	Multiple Sex Offence Types
14	Physical Harm to Victim(s) in Sex Offences
15	Uses of Weapons or Threats of Death in Sex Offences
16	Escalation in Frequency or Severity of Sex Offences
17	Extreme Minimization or Denial of Sex Offences
18	Attitudes that Support or Condone Sex Offences
Future Plans	
1	Lacks Realistic Plans
2	Negative Attitude Toward Intervention

Adapted from Boer, Hart, Kropp, and Webster (1997).

Appendix 1.6.

Scales and Items in the RSVP

Scale	Item
Sexual Violence History	
1	Chronicity of Sexual Violence
2	Diversity of Sexual Violence
3	Escalation of Sexual Violence
4	Physical Coercion in Sexual Violence
5	Psychological Coercion in Sexual Violence
Psychological Adjustment	
6	Extreme Minimization or Denial of Sexual Violence
7	Attitudes that Support or Condone Sexual Violence
8	Problems with Self-Awareness
9	Problems with Stress or Coping
10	Problems Resulting from Child Abuse
Mental Disorder	
11	Sexual Deviance
12	Psychopathic Personality Disorder
13	Major Mental Illness
14	Problems with Substance Use
15	Violent or Suicidal Ideation
Social Adjustment	
16	Problems with Intimate Relationships
17	Problems with Non-Intimate Relationships
18	Problems with Employment
19	Non-Sexual Criminality
Manageability	
20	Problems with Planning
21	Problems with Treatment
22	Problems with Supervision

Adapted from Hart et al. (2003).

Appendix 1.7.

Scales and Items in the SARA

Scale	Item
Criminal History	
1	Past Assault of Family Members
2	Past Assault of Strangers or Acquaintances
3	Past Violation of Conditional Release ir Community Supervision
Psychosocial Adjustment	
4	Recent Relationship Problems
5	Recent Employment Problems
6	Victim of and/or Witness to Family Violence as a Child or Adolescent
7	Recent Substance Abuse/Dependence
8	Recent Suicidal or Homicidal Ideation/Intent
9	Recent Psychotic and/or Manic Symptoms
10	Personality Disorder with Anger, Impulsivity, or Behavioral Instability
Spousal Assault History	
11	Past Physical Assault
12	Past Sexual Assault/Sexual Jealousy
13	Past Use of Weapons and/or Credible Threats of Death
14	Recent Escalation in Frequency or Severity of Assault
15	Past Violation of "No Contact" Orders
16	Extreme Minimization or Denial or Spousal Assault History
17	Attitudes that Support or Condince Spousal Assault
Alleged (Current) Offence	
18	Severe and/or Sexual Assault
19	Uses of Weapons or Threats of Death
20	Violation of "No Contact" Order

Adapted from Kropp, Hart, Webster, and Eaves (1995).

Appendix 1.8.

Scales and Items in the SAVRY

Scale	Item
Historical Risk Factors	
1	History of Violence
2	History of Non—Violent Offending
3	Early Initiation of Violence
4	Past Supervision/Intervention Failures
5	History of Self-Harm or Suicide Attempts
6	Exposure to Violence in the Home
7	Childhood History of Maltreatment
8	Parental/Caregiver Criminality
9	Early Caregiver Disruption
10	Poor School Achievement
Social/Contextual Risk Factors	
11	Peer Delinquency
12	Peer Rejection
13	Stress and Poor Coping
14	Poor Parental Management
15	Lack of Personal/Social Support
16	Community Disorganization
Individual Risk Factors	
17	Negative Attitudes
18	Risk Taking/Impulsivity
19	Substance Use Difficulties
20	Anger Management Problems
21	Low Empathy/ Remorse
22	Attention Deficit/Hyperactivity Difficulties
23	Poor Compliance
24	Low Interest/Commitment to School
Protective Factors	
P1	Prosocial Involvement
P2	Strong Social Support
P3	Strong Attachments and Bonds
P4	Positive Attitude Towards Intervention and Authority
P5	Strong Commitment to School
P6	Resilient Personality Traits

Adapted from Borum, Bartel, and Forth (2003).

Appendix 1.9.

Scales and Items in the ERASOR

Scale	Item
Sexual Interests, Attitudes, and Behaviours	
1	Deviant Sexual Interests (Younger Children, Violence, or Both)
2	Obsessive Sexual Interests/Preoccupation with Sexual Thoughts
3	Attitudes Supportive of Sexual Offending
4	Unwillingness to Alter Deviant Sexual Interests/Attitudes
Historical Sexual Assaults	
5	Ever Sexually Assaulted 2 or more Victims
6	Ever Sexually Assaulted Same Victim 2 or More Times
7	Prior Adult Sanctions for Sexual Assault(s)
8	Threats of, or Use of, Violence/Weapons During Sexual Offense
9	Ever Sexually Assaulted as a Child
10	Ever Sexually Assaulted a Stranger
11	Indiscriminate Choice of Victims
12	Ever Sexually Assaulted a Male Victim *(male offenders only)*
13	Diverse Sexual-Assault Behaviours
Psychosocial Functioning	
14	Antisocial Interpersonal Orientation
15	Lack of Intimate Peer Relationships/Social Isolation
16	Negative Peer Associations and Influences
17	Interpersonal Aggression
18	Recent Escalation in Anger or Negative Affect
19	Poor self-regulation of Affect and Behaviour (Impulsivity)
Family/Environmental Functioning	
20	High-Stress Family Environment
21	Problematic Parent-Offender Relationships/Parental Rejection
22	Parent(s) Not Supporting Sexual-Offense-Specific Assessment/Treatment
23	Environment Supporting Opportunities to Reoffend Sexually
Treatment	
24	No Development or Practice of Realistic Prevention Plans/Strategies
25	Incomplete Sexual-Offense-Specific Treatment

Adapted from Worling and Curwen (2001).

Appendix 1.10.

Scales and Items in the EARL-20B

Scale	Item
Family	
1	Household Circumstances
2	Caregiver Continuity
3	Supports
4	Stressors
5	Parenting Style
6	Antisocial Values and Conduct
Child	
1	Developmental Problems
2	Onset of Behavioral Difficulties
3	Trauma
4	Impulsivity
5	Likeability
6	Peer Socialization
7	School Functioning
8	Structured Community Act.
9	Police Contact
10	Antisocial Attitudes
11	Antisocial Behaviour
12	Coping Ability
Amenability	
1	Family Responsivity
2	Child Treatability

Adapted from Augimeri, Webster, Koegl, and Levene (1998).

Appendix 1.11.

Coding Booklet

CODING FORM

Miscellaneous notes:

Study Characteristics

1) Reference. Provide full citation for the dissemination.

2) Year
   a. Provide year of publication, dissemination, or copyright: _____
   b. Under review (and no date provided)
   c. Not specified/no date

3) Name of first author's affiliation: _____

4) Type of affiliation of first author
   a. Academic
   b. Government
   c. Correctional
   d. Hospital/Clinic
   e. Mixed
   f. Other: _____

5) Document type
   a. Book chapter
   b. Journal article (peer reviewed)
   c. Journal article (under review)
   d. Report, not peer reviewed
   e. Conference Presentation
   f. PhD Dissertation
   g. MA Thesis
   h. Other, specify:_____

Sample Characteristics/Demographics

6) Country. Country in which data were collected.

7) Number of samples analyzed in the dissemination: ____

8) Sample overlap. References of disseminations that overlap with sample in present dissemination.

9) Sample size *(used for analyses)*.

10) Gender *(used for analyses)*.
    a. All women
    b. All men
    c. Mixed gender
    d. Unclear/not specified.

11) Number of women/girls *(used for analyses)*.

12) Percentage of sample comprising women/girls *(used for analyses)*.

13) Number of men/boys *(used for analyses)*.

14) Percentage of sample comprising men/boys *(used for analyses)*.

15) Race/Ethnicity.
    a. Number of participants described as Black, African-Canadian, or African-American.
    b. Percentage of participants described as Black, African-Canadian, or African-American.
    c. Number of participants described as White, Caucasian, or Anglo.
    d. Percentage of participants described as White, Caucasian, or Anglo.
    e. Number of participants described as Aboriginal, Native Canadian, or Native American.
    f. Percentage of participants described as Aboriginal, Native Canadian, or Native American.
    g. Number of participants described as Hispanic or Latina/Latino.
    h. Percentage of participants described as Hispanic or Latina/Latino.
    i. Number of participants described as Asian or Pacific Islander.
    j. Percentage of participants described as Asian or Pacific Islander.
    k. Number of participants described as Mixed or Other.
    l. Percentage of participants described as Mixed or Other.

16) Mean age (in years).

17) Standard deviation of age (in years).

18) Low range of age (in years).

19) High range of age (in years).

20) Age category
   a. Adult = 18+
   b. Juvenile = 13-17
   c. Child = 12 and younger

21) Sample type1
   a. Civil psychiatric
   b. Forensic psychiatric
   c. Correctional
   d. School
   e. Emergency Department
   f. Mixed
   g. Other

22) Sample type2 *(additional descriptor if needed/applicable)*
   a. Sex offender
   b. Mentally disordered offender
   c. Bias motivated offender
   d. In treatment

23) Supervision. Were participants under any form of supervision?
   a. Yes
   b. No
   c. Not specified
   d. Other: _____

Method

24) Selection of participants
   a. Random
   b. Consecutive (admissions or discharges)
   c. All people in one place meeting criteria
   d. Matched
   e. Other: _____
   f. Not specified

25) Follow-up
   a. Fixed. All participants were followed-up for exactly the same length of time (e.g., no differences in standard deviation, and min/max length variables below will be the same length of time)
   b. Variable. Participants were follow-up for different lengths of time (e.g., criminal conviction data recorded on the same date for all participants, but they had different discharge dates).

26) Average length of follow up (in days)

27) Standard deviation of mean length of follow up (in days)

28) Minimum length of follow-up (in days)

29) Maximum length of follow-up (in days)

30) Setting
   a. Institutional (prison, hospital)
   b. Community
   c. Institutional + Community

31) Design of study
   a. True prospective.
      - Code risk measure in present day, real time; record violence data at some point in the future.
   b. Pseudo-prospective (aka 'retrospective follow-up').
      - Violence has already occurred in real time. Code risk measure using only information that was available prior to violence happening or record score for a risk measure from a file/report that was made at some point in the past but before violence had occurred; then, record violence at some point in time after the risk judgment was made (but, again, in real time, the violence has already happened).
   c. Unclear whether true or pseudo-prospective
   d. Retrospective.
      - Risk judgment is made, and the study looks at violence that occurred before that judgment was made (e.g., correlations with a score and past charges).

Information about the Risk Tool/Independent Variable(s)

32) Number of risk tools studied

33) Name of risk tools studied

34) Version of risk tool:
   a. Not specified

35) If HCR-20 is scored:
   a. Risk Management scale scored as IN
   b. Risk Management scale scored as OUT
   c. Not specified

36) Item selection
   a. Statistical
   b. Logical/rational
   c. Unclear

37) Item Integration
   a. Statistical
   b. Rational/SPJ

38) Format Risk Tool
   a. Assessor rated
   b. Self-report
   c. Mixed, self-report and assessor
   d. Unclear

39) Allegiance
   a. Yes, one of the co-authors of the dissemination is an author of the risk tool
   b. Yes, one of the co-authors of the dissemination is/was a graduate student of an author of the risk tool
   c. No
   d. Unclear

40) Mean score of risk index

41) Standard deviation of risk index

42) Minimum value of risk score

43) Maximum value of risk score

44) Were any items scored with replacement/modification?
   a. Yes
   b. No

45) Sources used to score risk tool
   a. Interview only
   b. Files + Interview
   c. Files only (no additional information)
   d. Files only, and interview not part of scoring
   e. Files only, but included reports/evaluations based on previous interview
   f. Risk Score recorded from file
   g. Not applicable (self-report tool)
   h. Not specified
   i. Other: _____

Information about the Outcome/Dependent Independent Variable(s)

46) Operationalization. Quote the definition of outcome/violence provided.

47) Page number operationalization. Number of page in dissemination where outcome is defined.

48) Outcome category
   a. Any negative outcome/reconviction/charge, etc. (it is either a mish-mash of other categories or the outcome is described as 'any recidivism')
   b. Violence (outcome is described as 'violent' but there are not enough details to tell whether it was only physical violence)
   c. Physical violence (specified that outcome was actual or attempted physical violence to another person, or a threat made with a weapon in hand)
      i. Physical violence - specifically states <u>includes</u> sexual violence
      ii. Physical violence - specifically states <u>excludes</u> sexual violence
   d. Sexual
   e. Nonviolent (nonphysical, nonsexual outcomes such as fraud, failure to appear, drug offences, theft, etc.)
   f. Destruction to property
   g. Verbal abuse/ aggression/threats (not further details, and not specified if made with a weapon in hand)
   h. Destruction of property + Verbal abuse//aggression/ threats

49) Base rates of each outcome

50) Was the base rate set artificially by known groups sampling?

a. Yes
b. No

51) Sources used to assess outcome status

a. Official criminal justice/legal records
b. Other official files (e.g., hospital records, Department of Motor Vehicle records, etc.)
c. Self-report
d. Collateral source
e. Staff Observation
f. Other: _____
g. Not specified

52) Number of sources used to determine outcome status.

a. Specify: _____
b. Unclear

53) Type of sanction:

a. Arrest
b. Charge
c. Conviction
d. Revocation
e. Return to institution
f. Not specified

Information about raters and rating procedure

54) Raters' education (code all that apply)

a. Undergraduate
b. Masters
c. Doctoral
d. M.D.
e. Other
f. Mixed
g. Not specified

55) Raters' discipline (code all that apply)

a. Research
b. Nursing
c. Social Work
d. Psychology
e. Psychiatry
f. Criminology
g. Parole Board
h. Parole/Probation Officers
i. Clinician
j. Multidisciplinary treatment team
k. Mixed
l. Self-report
m. Not specified

56) Raters' training. Were the raters trained?

a. No
b. Yes
c. Yes, training to a preset level of reliability provided
d. Not applicable (self report measure)
e. Not specified

57) Number of raters used in the study

a. Specify the number of coders in the study: ____
b. Not specified

58) Number of raters who coded data for a single participant.

a. Specify number: ____
b. N/A (self report)

59) If 2+ raters coded data for each participant, scores used for analyses:

a. Consensus scores
b. Averaged scores
c. Other: ____
d. Not specified

60) Blind. Was the rater of the risk tool blind to outcome status?

a. Yes
b. No

c. Not applicable (true prospective, or risk score recorded from file in pseudo-prospective design)
d. Not specified

61) Purpose of risk assessment
   a. Clinical practice/decision-making
   b. Research only

62) Prediction/assessment range. Was the length of time for which prediction/assessment was being made specified?
   a. Length of time (in days):
   b. Not specified

63) Interrater reliability
   a. IRR data presented?
      i. Yes
      ii. No

   b. IRR Type
      i. ICC1
      ii. ICC2
      iii. Correlation
      iv. Kendalls tau
      v. Kappa
      vi. Other: ____
      vii. Not specified

   c. Number of cases on which IRR is based: _____ (not specified)

   d. Specify value of reliability coefficient: _____ (not specified)

   e. Low range of 95% Confidence Interval: ____ (not specified)

   f. High range of 95% Confidence Interval: ____ (not specified)

Effect Sizes

64) Number of effect sizes that are codeable in the study: _____

65) Type of effect size
   a. *r*
   b. *F*
   c. *t*
   d. AUC
   e. OR
   f. *d*
   g. *d* calculated from mean (SD)
   h. Chi-square
   i. Relative Risk
   j. Standardized beta
   k. Other:

66) Original value of effect size

67) Number assigned to effect size within this dissemination

68) Value of AUC (estimated from original value if AUC not provided)

69) Standard error, AUC

70) Low range of 95% Confidence Interval, AUC:

71) High range of 95% Confidence Interval, AUC:

72) Page number in dissemination where effect sizes are presented:

73) Categories. Is additional information on predictive validity of risk categories presented?
   a. Yes
   b. No

Incremental validity analyses

Within-study comparisons between SPJ risk tool and non-SPJ tools

74) Are data presented regarding incremental validity relative to other risk assessment tools?
   a. Page number(s) where data are presented
   b. Names of risk tools being compared
   c. Detail results:

Within-study comparisons between SPJ risk tool and putative risk factors

75) Are data presented regarding incremental validity relative to other putative risk factors?
   a. Page number(s) where data are presented
   b. Names of other variables in model being compared
   c. Detail results:

Effect Size (ES) Summary Table

ES #	Page #	ES value (with SD, 95% CI, etc. if presented)	Outcome category (#48)	Risk index (total or subscale)	Additional moderator info if necessary to clarify (gender, setting, etc)
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					

***Additional Summary Table on page of coding book if needed

Appendix 1.12.

Transformational Formulae

1. **From *r*:**

From Rice and Harris (2005); after Rosenthal (1991) and Swets (1986):

$$d = \frac{r}{\sqrt{pq\,(1 - r^2)}}$$

p = the base rate; $q = 1 - p$,

$$d = \sqrt{2} \cdot z(AUX),$$

where z(AUC) is the z-transform or normal deviate of AUC, and where variances of the two populations are equal.

2. **From *d*:**

From Rice and Harris (2005); after Rosenthal (1991) and Swets (1986):

$$r = \frac{d}{\sqrt{d^2 + \left(\frac{1}{pq}\right)}}$$

$$d = \frac{r}{\sqrt{pq\,(1 - r^2)}}$$

p = the base rate; $q = 1 - p$,

$$d = \sqrt{2} \cdot z(AUX),$$

where z(AUC) is the z-transform or normal deviate of AUC, and where variances of the two populations are equal.

If *d* needed to be estimated because the article presented only group means and standard deviations, formula 1 presented in Table B10 by Lipsey and Wilson (2001, p. 198) was used:

$$d = (\bar{X}_1 - \bar{X}_2) / \sqrt{s_{pooled}}$$

$$s_{pooled} = \sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 / n_1 + n_2 - 2}$$

3. **From OR**:

d was estimated using formula 6 presented in Table B12 by Lipsey and Wilson (2001, p. 202):

$$d = \frac{\ln(OR)\sqrt{3}}{\pi}$$

Next, the AUC was estimated using the formulae presented above.

If an odds ratio needed to be estimated because an article presented only frequencies, the following formula from Hasselblad and Hedges (1995) was used:

OR = ad/bc, where a, b, c, and d are:

	Above cut score	Below cut score
Violent	a	b
Not violent	c	d

4. **From χ^2**:

A chi square value (χ^2) with $df = 1$ and total sample size (N) was converted to r as follows (Lipsey & Wilson, 2001, p. 201):

$$|r| = \sqrt{\frac{\chi^2}{N}}$$

r subsequently was converted to an AUC using the formulae presented above.