

# International Perspectives on Forensic Risk Assessment: Measuring Use, Perceived Utility, and Research Quality

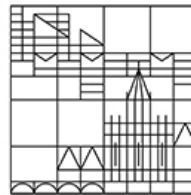
Dissertation zur Erlangung des akademischen Grades des Doktors der Naturwissenschaften  
(Dr. rer. nat.)

vorgelegt von

**Jay P. Singh**

an der

Universität  
Konstanz



Mathematisch-Naturwissenschaftliche Sektion  
Fachbereich Psychologie

Tag der mündlichen Prüfung: 18. Januar 2016

1. Referent: apl. Prof. Dr. Jérôme Endrass
2. Referentin: PD. Dr. Astrid Rossegger

# TABLE OF CONTENT

---

SUMMARY .....	VIII
ZUSAMMENFASSUNG .....	IX
INTRODUCTION .....	1
1.1 Identifying Risk and Protective Factors .....	1
1.2 Contemporary Approaches to Forensic Risk Assessment .....	3
1.2.1 Unstructured Clinical Judgment.....	3
1.2.2 Actuarial Assessment.....	5
1.2.3 Structured Professional Judgment .....	12
1.3 Integrating Risk Assessment Research into Practice .....	19
1.4 The Present Thesis .....	20
2 RESEARCH ARTICLES.....	25
2.1 Hurducas, C., Singh, J. P., de Ruiter, C., & Petrila, J. (2014). Violence risk assessment tools: A systematic review of clinical surveys. <i>International Journal of Forensic Mental Health, 13</i> , 181-192. ....	25
2.1.1 Abstract.....	25
2.1.2 Introduction .....	26
2.1.3 Method .....	28
2.1.4 Results .....	32
2.1.5 Discussion.....	42
2.2 Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., de Ruiter, C., Dean, K., Doyle, M., Folino, J., Godoy, V., Grann, M., Ho, R., Large, M., Nielsen, L. H., Pham, T., Rebocho, R. M., Reeves, K., Rettenberger, M., & Otto, R. K. (2014). Use and perceived utility of structured violence risk assessment tools in 44 countries: Findings from the IRiS Project. <i>International Journal of Forensic Mental Health, 13</i> , 193-206. ....	46
2.2.1 Abstract.....	46
2.2.2 Introduction .....	47
2.2.3 Methods .....	49
2.2.4 Results.....	54
2.2.5 Discussion.....	60
2.2.6 Funding.....	64
2.2.8 Appendix.....	65
2.3 Singh, J. P., Desmarais, S., & Van Dorn, R. A. (2013). Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. <i>Behavioral Sciences &amp; the Law, 31</i> , 55-73.....	68
2.3.1 Abstract.....	68
2.3.2 Introduction .....	68
2.3.3 Methods .....	73
2.3.4 Results.....	77
2.3.5 Discussion.....	84
2.3.6 Conclusion.....	87

2.3.7	Acknowledgements .....	88
2.3.8	Appendix .....	88
<b>2.4</b>	<b>Singh, J. P., Yang, S., Mulvey, E., &amp; the RAGEE Group (2014). Reporting standards for risk assessment predictive validity studies: The Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement. <i>Law &amp; Human Behavior</i>, 39, 15-22.....</b>	<b>93</b>
2.4.1	Abstract.....	93
2.4.2	Introduction .....	93
2.4.3	Method .....	95
	.....	99
2.4.4	Results .....	99
2.4.5	Discussion.....	102
2.4.6	Conclusion.....	104
<b>2.5</b>	<b>Gmür, C., &amp; Singh, J. P. (2014). Reporting quality of the sex offender risk assessment literature: A systematic review. <i>Sexual Offender Treatment</i>, 9, 1-14.....</b>	<b>105</b>
2.5.1	Abstract.....	105
2.5.2	Introduction .....	105
2.5.3	Method .....	107
2.5.4	Results.....	109
2.5.5	Discussion.....	110
2.5.6	Conflict of Interest and Funding .....	113
	.....	115
2.5.7	Supplement.....	117
 <b>RECORD OF ACHIEVEMENT .....</b>		 <b>119</b>
 <b>REFERENCES.....</b>		 <b>121</b>
 <b>RAGEE SUPPLEMENT .....</b>		 <b>161</b>

## LIST OF TABLES

---

<b>Table 1:</b> Characteristics of Commonly-Used Risk Assessment Tools .....	22
<b>Table 2:</b> Types of Items Included in Commonly-Used Risk Assessment Tools .....	23
<b>Table 3:</b> Item Content Included in Commonly-Used Risk Assessment Tools .....	24
<b>Table 4:</b> Methodological Quality Markers in Nine Clinical Surveys of Use of Structured Violence Risk Assessment Instruments .....	32
<b>Table 5:</b> Characteristics Concerning the Administration of Nine Clinical Surveys of Structured Violence Risk Assessment .....	40
<b>Table 6:</b> Characteristics Concerning the Findings of Nine Clinical Surveys on Structured Violence Risk Assessment.....	41
<b>Table 7:</b> Demographic and Clinical Characteristics of Survey Respondents by Continent.....	51
<b>Table 8:</b> Demographic and Clinical Characteristics of Survey Respondents by Professional Discipline .....	52
<b>Table 9:</b> Risk Assessment Instrument Prevalence and Frequency of Use Over the Past 12 Months by Continent .....	55
<b>Table 10:</b> Risk Assessment Instrument Prevalence and Frequency of Use Over the Past 12 Months by Professional Discipline.....	56
<b>Table 11:</b> Instrument Use in Violence Risk Assessment, Management, and Monitoring Over the Past 12 Months.....	59
<b>Table 12:</b> Characteristics of the 25 risk assessment tools included in the review .....	74
<b>Table 13:</b> Reporting characteristics of studies investigating the predictive validity of 25 risk assessment tools .....	79
<b>Table 14:</b> Characteristics of the 25 risk assessment tools included in the review .....	98
<b>Table 15:</b> Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement Checklist .....	101
<b>Table 16:</b> Reporting quality of two leading sex offender risk assessment journals according to Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement checklist criteria .....	114

## LIST OF FIGURES

---

<b>Figure 1:</b> Systematic search for surveys investigating violence risk assessment tool use and perceived utility. ....	29
<b>Figure 2:</b> Countries participating in an international survey on violence risk assessment practices. ...	50
<b>Figure 3:</b> Variation in magnitude benchmarks for the area under the curve (AUC) in structured risk assessment tools. ....	82
<b>Figure 4:</b> Identification of methodological reporting checklists for prognostic and diagnostic accuracy studies. ....	99
<b>Figure 5:</b> Development of the Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) statement. ....	100

## ABBREVIATIONS

---

<b>AUC</b>	Area Under the Curve
<b>M</b>	Mean
<b>SD</b>	Standard Deviation
<b>PRISMA</b>	Preferred Reporting Items for Systematic Reviews and Meta-analyses
<b>N</b>	Number of Participants
<b>K</b>	Number of Samples
<b>RAGEE</b>	Risk Assessment Guidelines for the Evaluation of Efficacy
<b>IRiS</b>	International Risk Survey
<b>AP-LS</b>	American Psychology-Law Society
<b>ABFP</b>	American Board of Forensic Psychology
<b>MBP</b>	Michigan Board of Psychology
<b>IAFMHS</b>	International Association of Forensic Mental Health Services
<b>CFMHS</b>	Community Forensic Mental Health Services
<b>CPA-CJS</b>	Canadian Psychological Association – Criminal Justice Section
<b>ACFP</b>	American College of Forensic Psychology
<b>COVR</b>	Classification of Violence Risk
<b>FOTRES</b>	Forensisches Operationalisiertes Therapie-Risiko-Evaluations System
<b>HCR-20</b>	Historical, Clinical, Risk Management-20
<b>HKT-30</b>	Historische, Klinische, Toekomstige-30
<b>LSI-R</b>	Level of Service Inventory-Revised
<b>PCL-R</b>	Psychology Checklist-Revised
<b>PCL:SV</b>	Psychopathy Checklist: Screening Version
<b>SAPROF</b>	Structured Assessment of Protective Factors

<b>START</b>	Short-Term Assessment of Risk and Treatability
<b>V-RISK-10</b>	Violence Risk Screening-10
<b>VRAG</b>	Violence Risk Appraisal Guide
<b>VRS</b>	Violence Risk Scale
<b>CI</b>	Confidence Interval
<b>SPJ</b>	Structured Professional Judgment
<b>UCJ</b>	Unstructured Clinical Judgment
<b>PPV</b>	Positive Predictive Value
<b>NPV</b>	Negative Predictive Value
<b>NND</b>	Number Needed to Detain
<b>NSD</b>	Number Safely Discharged
<b>ROC</b>	Receiver Operating Characteristics
<b>GSIR</b>	General Statistical Information on Recidivism
<b>LS/CMI</b>	Level of Service/Case Management Inventory
<b>MnSOST-R</b>	Minnesota Sex Offender Screening Tool-Revised
<b>RRASOR</b>	Rapid Risk Assessment for Sexual Offense Recidivism
<b>VORAS</b>	Violent Offender Risk Assessment Scale
<b>SORM</b>	Structured Outcome Assessment and Community Risk Monitoring
<b>SARA</b>	Spousal Assault Risk Assessment
<b>CHERRIES</b>	Checklist for Reporting Results of Internet E-Surveys
<b>EQUATOR</b>	Enhancing the Quality and Transparency of Health Research
<b>SAJRT</b>	Sexual Abuse: A Journal of Research and Treatment
<b>JSA</b>	Journal of Sexual Aggression
<b>NHST</b>	Null Hypothesis Significance Testing
<b>OGRS</b>	Offender Group Reconviction Scale

## SUMMARY

---

Forensic risk assessment refers to the attempt to predict the likelihood of future violence in order to identify individuals in need of intervention. Risk assessment protocols have been implemented in mental health and criminal justice settings around the globe to prioritize risk reduction strategies for those most at need. Helping to allocate scarce resources more effectively and efficiently while protecting our communities, risk assessment has come to be a cornerstone of forensic practice in many jurisdictions. The present thesis investigates the use, perceived utility, and research quality on forensic risk assessment tools. First, a systematic review of qualitative surveys on forensic risk assessment tools is conducted to explore the international use and perceived utility of such instruments (Article 1). The systematic review concludes that there is a lack of methodologically-rigorous qualitative research on forensic risk assessment tool use and perceived utility across continents and professional disciplines, necessitating a new, large-scale study (Article 2). Given the lack of methodologically-rigorous qualitative research on forensic risk assessment tools, a second systematic review is conducted to explore the methodological rigor of the quantitative research literature concerning the predictive validity of these instruments (Article 3). The systematic review concludes that the field could benefit from the development of standardized guidelines promoting the transparent and consistent reporting of methodology and results in peer-reviewed publications on violence risk assessment tools (Article 4). These standardized guidelines are then applied to the sexual offender risk assessment tool research literature to examine its transparency and consistency of reporting of methodology and results (Article 5).



## ZUSAMMENFASSUNG

---

Forensische Risikobeurteilung beschreibt den Versuch, die Wahrscheinlichkeit von zukünftiger Gewalt vorherzusagen, um Personen zu identifizieren, bei denen Interventionen möglich und zielführend sind. Beurteilungsprotokolle für Risikoeinschätzungen wurden weltweit sowohl in psychologischen als auch psychiatrischen Settings, aber auch im juristischen und forensischen Kontext eingeführt, um Reduktionsstrategien für Täter zu priorisieren, die in dieser Hinsicht die grössten Bedürfnisse aufweisen. Risikobeurteilungen wurden folglich auch zu einem zentralen Thema in der forensischen Praxis, um die wenigen verfügbaren Ressourcen effektiver und effizienter zu nutzen und gleichzeitig den gesellschaftlichen Schutz zu garantieren. Die vorliegende Arbeit untersucht daher die Benutzung, die wahrgenommene Nützlichkeit und die Qualität der Forschungsarbeiten rund um forensische Risikobeurteilungsmethoden. Zuerst wird eine systematische Übersichtsarbeit zu qualitativen Umfragen über die Benutzung forensischer Risk-Assessment Instrumente berichtet, die die internationale Anwendung und den wahrgenommenen Nutzen dieser Instrumente zum Thema haben (Artikel 1). Die folgende Übersichtsarbeit kommt zu dem Schluss, dass bezüglich methodisch hochwertiger Forschung zu forensischen Riskobeurteilungsmitteln, sowie zu deren wahrgenommener Nützlichkeit ein deutlicher Mangel besteht, der über Kontinente und Professionen hin variiert. Hinsichtlich dessen besteht Bedarf nach weiteren umfassenden Studien (Artikel 2). In Anbetracht dieses Mangels wurde eine weitere Übersichtsarbeit durchgeführt, die die methodische Qualität bisheriger quantitativer Forschungsarbeiten zur Vorhersagevalidität forensischer Risikobeurteilungen untersucht (Artikel 3). Dass das forensische Umfeld im Rahmen des Risk-Assessments für gewalttätige Vorfälle von Vorgaben für eine transparente und konsistente Kommunikation über methodische Vorgehensweisen und Ergebnisse in peer-reviewten Publikationen

profitiert, zeigt die darauf folgende systematische Übersichtsarbeit (Artikel 4). Diese standardisierten Vorgaben können schliesslich auch auf die Forschungsliteratur zu Risk-Assessment Instrumenten bei Sexualstraftätern Anwendung finden, mit dem Ziel, Transparenz und Konsistenz in berichteter Methodik und Resultaten zu berichten (Artikel 5). Die Ergebnisse der vorliegenden Arbeit zeigen sich als wegweisend für sowohl Forschung als auch für die forensische Praxis und (juristische) Entscheidungsträger.

# INTRODUCTION

---

Historically, the construct of *risk* has referred to the probability of gain or loss weighted by the value of what stands to be gained or lost. Beginning largely in the 20<sup>th</sup> century, this construct was applied to the area of forensic mental health. *Forensic risk assessment* refers to the process by which the likelihood of future antisocial behavior is evaluated (Singh, 2012). The antisocial behavior being predicted may constitute a first-time offense or a repeat offense, the latter of which is referred to as *recidivism*. Risk assessments routinely involve the structured examination of a number of *risk factors* (biological, psychological, or sociological characteristics that increase the likelihood of antisocial behavior) and *protective factors* (biological, psychological, or sociological characteristics that decrease the likelihood of antisocial behavior). These may be either *static* (historical or unchanging), *acutely dynamic* (modifiable and likely to change), or *stably dynamic* (modifiable but unlikely to change) in nature (Andrews & Bonta, 2010). An example of a static risk factor for antisocial behavior is a history of violence (Quinsey, Harris, Rice, & Cormier, 2006), whereas an acute dynamic risk factor would be stress (Borum, 1996), and a stable dynamic risk factor would be marital status (Andrews & Bonta, 1995). An example of a static protective factor against antisocial behavior is intelligence (de Vogel, de Ruiter, Bouman, & de Vries Robbe, 2007), whereas an acute dynamic protective factor would be medication adherence (Webster, Martin, Brink, Nicholls, & Desmarais, 2009), and a stable dynamic protective factor would be healthy peer relationships (Webster et al., 2009).

## 1.1 Identifying Risk and Protective Factors

According to the guidelines set forth by Grann and Långström (2007), risk and protective factors – be they static or dynamic in nature – can be identified using one of three

techniques: (a) the empirical method, (b) the theoretical method, or (c) the clinical method. Each of these three techniques has its own merit, albeit they vary in terms of their focus on psychometrics versus practical application. In the empirical method, risk and protective factors are identified through research in which a sample is followed for such a duration as to allow for the possibility of offending. The biopsychosocial characteristics of those who offend are analyzed to see if they systematically differ from those who do not. If the presence of a given characteristic increases the likelihood of offending to a statistically significant extent, it is considered a risk factor. If the presence of that characteristic decreases this likelihood, the characteristic is considered a protective factor.

In the theoretical method, a particular theory (e.g., psychoanalytic, behavioral, cognitive) is used to guide decisions as to which characteristics place an individual at a higher or lower risk of antisocial behavior (Grann & Långström, 2007). Different theoretical orientations offer different conceptualizations of what constitutes an “at risk” person and propose different mechanisms concerning how that individual came to be at risk. For example, risk assessments formulated from a psychoanalytic perspective may take into consideration information concerning disorganized attachment styles as well as an individual’s sexual history. (For an overview of forensic risk assessment and psychodynamic theory, see Doctor & Nettleton, 2003). Behavioral measures, on the other hand, would be more likely to include consideration of the individual’s previous offending history, social competence, and his or her parents’ style of discipline. (For an overview of forensic risk assessment and behavioral theory, see Eifert & Feldner, 2004). Risk tools adopting a cognitive approach would likely include consideration of an individual’s capacity for emotion regulation, tendency to ruminate, and level of impulsivity. (For an overview of cognitive approaches to forensic risk management, see Lipsey, Hapman, & Landenberger, 2001).

In contrast to the previous two approaches, the clinical method of identifying risk and protective factors involves identifying individual characteristics which, regardless of whether they are empirically or theoretically associated with offending, are changeable and thus can be addressed through clinical intervention (Grann & Långström, 2007). For example, although traits such as an individual's history of antisocial behavior or severe mental illness cannot be altered, other characteristics such as employment status or level of education can. Hence, the clinical method places an emphasis on dynamic factors.

## 1.2 Contemporary Approaches to Forensic Risk Assessment

Although there are numerous adverse outcomes that can be evaluated through forensic risk assessment (e.g., substance use, absconsion, self-harm), the current article will focus on evidence-based approaches to violence, sex offender, and general recidivism risk assessment. Specifically, we will explore the three leading approaches to risk assessment currently used in practice and examples of key tools that follow each. In addition, we will examine the importance of understanding Tarasoff liability in the context of forensic risk assessment.

Systematic reviews and meta-analyses of the research base on forensic risk assessment have established three leading approaches to this form of evaluation: (a) unstructured clinical judgment, (b) actuarial assessment, and (c) structured professional judgment (Singh & Fazel, 2010). In the following section, we will examine the relative strengths and weaknesses of each.

### 1.2.1 Unstructured Clinical Judgment

Unstructured clinical judgment (UCJ) refers to the subjective process of evaluating the likelihood of an adverse outcome without the use of a structured method (e.g., a risk assessment tool). Instead, clinical skills and experience with the given individual whose risk is being assessed are relied upon (Murray & Thomson, 2010). The key benefits of the UCJ

approach include its flexibility, its utility in tailoring risk assessments to a given individual, its incorporation of a variety of case-specific risk and protective static and dynamic factors, and its inexpensiveness (i.e., no materials need to be purchased). The key drawback of the UCJ approach is its inherent subjectivity, resulting in poor rates of reliability and predictive validity. Of particular concern is this approach's vulnerability to human judgment biases in the decision-making process. For example, hindsight bias due to recent tragic events involving high-profile homicides by individuals diagnosed with a mental illness may result in the overestimation of violence-risk in persons with quite low base rates of interpersonal aggression (Arkes, 1991; Large, Ryan, Singh, Paton, & Nielssen, 2011). If evaluating a college-aged adolescent in Newtown, Connecticut in the United States, who was diagnosed with Asperger's Syndrome and raised by a single mother who had taught him to fire guns, this adolescent would likely be perceived as a higher risk immediately after an armed gunman reportedly diagnosed with Asperger's Syndrome entered Sandy Hook Elementary School in Newtown in 2012 and fatally shot 20 children and six adult staff members. This despite epidemiological research findings suggesting that individuals diagnosed with Asperger's Syndrome are not at increased risk of violence compared to the general population (Ghaziuddin, 2013) and that the large majority of individuals with this diagnosis who do go on to be violent do not commit crimes involving weapons (Harmon, 2012).

Perhaps the best-known criticism of unstructured clinical judgment in forensic risk assessment is the seminal monograph by Monahan (1981), entitled *The Clinical Prediction of Violent Behavior*. A spiritual successor to Meehl's (1954) *Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, in which it was argued that professionals cannot predict outcomes as successfully as statistical formulae, Monahan reviewed the research literature on unstructured clinical judgment and found that clinicians are unable to predict violence at rates above chance, concluding:

[P]sychiatrists and psychologists are accurate in no more than one out of three predictions of violent behavior over a several-year period among institutionalized populations that had both committed violence in the past (and thus had high base rates for it) and those who were diagnosed as ‘mentally ill’ (Monahan, 1981, p. 48-49).

### 1.2.2 Actuarial Assessment

Actuarial risk assessment tools are structured instruments composed of risk and/or protective, static and/or dynamic factors that are found to be associated with the adverse event of interest using a statistical methodology (e.g., logistic regression, Cox regression, Chi-Squared Automatic Interaction Detection [CHAID]). Each item is weighted in accordance with the amount of variance it accounts for in the prediction of the adverse event of interest. Total scores are cross-referenced with a manual in which estimates of recidivism rates are provided for either each score or for ranges of scores (referred to as “risk bins” or “risk categories”). These estimates are derived from the actual rates of recidivism seen in groups with the same score or ranges of scores in the sample on which the tool was calibrated (i.e., the group whose data was used to develop the tool).

The key benefits of actuarial risk assessment tools include their objectivity and transparency in the risk assessment process, their speed of administration, their requiring mostly historical information (i.e., incorporating mostly static risk factors) that are routinely available in criminal/court/medical records, their removal of human judgment biases inherent in the clinical decision-making process, and the generation of an estimated recidivism rate. The latter is perceived as the most significant strength of actuarial risk assessment tools, making them of higher perceived usefulness in legal settings (Singh, 2013).

The key drawbacks of actuarial risk assessment tools are the inability to apply group-based recidivism rates to individual patients (Hart, Michie, & Cook, 2007), the instability of estimated recidivism rates when applied to groups in different jurisdictions (Singh, Fazel,

Gueorguieva, & Buchanan, 2014), and the inability to incorporate case-specific information to modify estimated recidivism rates. Concerning the latter, the preponderance of the research literature on modifying the findings of actuarial risk assessment tools suggests that such modification weakens rather than strengthens their reliability and predictive validity (Quinsey et al., 2006). In addition, adding or removing additional items on actuarial risk assessment tools or using them with unintended populations or to predict unintended outcomes has been found to weaken their predictive validity (Quinsey et al., 2006).

Examples of commonly-used actuarial assessment schemes include the Classification of Violence Risk (COVR; Monahan et al., 2005); Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995); Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004); Psychopathy Checklist-Revised (PCL-R; Hare, 2003); Psychopathy Checklist: Screening Version (PCL:SV; Hart, Cox, & Hare, 1995); Risk Matrix 2000 (RM2000; Thornton et al., 2003); Static-99 and Static-99R (Hanson & Thornton, 1999; Helmus, Thornton, Hanson, & Babchishin, 2012); Violence Risk Appraisal Guide (VRAG; Quinsey, Harris, Rice, & Cormier, 2006); Sex Offender Risk Appraisal Guide (SORAG; Quinsey et al., 2006); and Violence Risk Scale (VRS; Wong & Gordon, 2009).

**Classification of Violence Risk (COVR).** The COVR is a computer-based actuarial instrument designed to predict the likelihood of violence in adult psychiatric inpatients upon discharge to the community (Monahan et al., 2005). The measure guides clinicians through a structured chart review and a 5-10 minute interview with the examinee in order to assess violence risk among individuals with a mental disorder in acute psychiatric care. Information gathered via self-report is used to score individuals on up to 106 items. Once the COVR assessment is completed, the software generates a report that contains a statistical estimate of the patient's violence risk, including the confidence interval for that estimate and a list of the risk factors that the program took into account to produce the estimate. Total scores are used



to classify individuals into one of five risk categories, each of which has an estimated recidivism rate. Systematic review evidence suggests that COVR assessments produce moderate levels of predictive validity when used to predict outpatient violence in both male and female psychiatric populations (Singh, Serper, Reinharth, & Fazel, 2011).

**Level of Service Inventory-Revised (LSI-R).** The LSI-R is an actuarial instrument designed to predict the likelihood of general recidivism in adult offenders according to the Risk-Need-Responsivity Model of (Andrews & Bonta, 1995). Information gathered via file review and a semi-structured interview is used to score individuals on 54 items measuring 10 domains: criminal history (10 items), leisure/recreation (2 items), education/employment (10 items), companions (5 items), financial (2 items), alcohol/drug problems (9 items), family/marital (4 items), emotional/personal (5 items), accommodation (3 items), and attitude/orientation (4 items). Total scores on the LSI-R are used to classify individuals into one of five risk categories (low, low-moderate, moderate, moderate-high, high), each of which has an estimated recidivism rate. The LSI-R has been found to be effective in predicting parole outcomes, success in halfway houses, institutional misconduct, and recidivism in the community (Andrews & Bonta, 2010).

**Level of Service/Case Management Inventory (LS/CMI).** The LS/CMI is an actuarial instrument designed to predict the likelihood of general recidivism in adult offenders (Andrews, Bonta, & Wormith, 2004). Information gathered via file review and semi-structured interview is used to score individuals on 43 items measuring eight domains, seven of which are derived from the LSI-R: criminal history (8 items), leisure/recreation (2 items), alcohol/drug problems (8 items), education/employment (9 items), companions (4 items), procriminal attitudes/orientation (4 items), family/marital (4 items), and antisocial patterns (4 items). Each of the eight domains also includes the opportunity to rate the domain as a strength that can be built upon in a case management plan. Strengths were introduced to the

scheme after LSI-R users commented that they preferred considering the totality of the offender being assessed when creating a management plan (Andrews et al., 2010). Total scores on the LS/CMI are used to classify individuals into one of five risk categories (very low, low, medium, high, or very high). The LS/CMI has been found to produce valid and reliable assessments of general recidivism risk across sexes, ethnicities, and Western countries (Olver, Stockdale, & Wormith, 2014).

**Psychopathy Checklist-Revised (PCL-R).** The PCL-R is a personality scale designed to diagnose psychopathy in adult forensic populations (Hare, 2003). The item content on the PCL-R is based on Cleckley's (1941) operational definition of the psychopathic personality outlined in his work, *The Mask of Sanity*. Information gathered via file review, semi-structured interviews, and collateral information is used to score individuals on 20 items that load onto two factors: interpersonal and affective features (10 items) and impulsive and antisocial behavior (10 items). Different cut-off scores have been put forth for use in different geographic regions and different populations. The PCL-R should be completed by a mental health professional (such as a psychologist or other professional trained in the field of mental health, psychology, or psychiatry) with extensive training (e.g., psychopathology, psychopathy, psychometrics) (Hare, 2003). Although the PCL-R was not originally designed as a risk assessment tool, meta-analytic reviews have found that the measure is able to predict future offending accurately in a number of settings and populations (Leistico, Salekin, DeCoster & Rogers, 2008; Salekin, Rogers & Sewell, 1996; Walters, 2003). It is also notable that PCL-R scores were originally incorporated into several risk assessment tools such as the HCR-20, SORAG, SVR-20, and VRAG.

**Psychopathy Checklist: Screening Version (PCL:SV).** The PCL:SV is a personality scale designed to screen for psychopathy in psychiatric inpatients (Hart, Cox, & Hare, 1995). It is an abbreviated and highly correlated version of the PCL-R, originally developed as part

of the MacArthur Violence Risk Assessment Study (Monahan et al., 2001). Analogous to the PCL-R, information gathered via file review and a semi-structured interview is used to score individuals on 12 items that load onto two factors: interpersonal and affective features (6 items), and impulsive and antisocial behaviors (6 items). Total scores on the PCL:SV are used to classify individuals as either non-psychopathic, psychopathic, or as needing a full PCL-R assessment before a diagnosis can be made. In the case of insufficient information, one item on each subscale can be prorated (Hart et al., 1995). The PCL:SV can be completed by assessors who have completed graduate-level courses in test administration and measurement, or with equivalent training. In addition, administrators must have an advanced degree in a relevant profession such as a field of mental health (Multi-Health Systems, 2015). As with the full PCL-R, although the PCL:SV was not designed as a risk assessment tool, meta-analytic research suggests that assessments using this measure are able to predict violent and general recidivism in a number of populations (Leistico et al., 2008).

**Risk Matrix 2000 (RM2000).** The RM2000 is an actuarial instrument designed to predict the likelihood of future sexual and non-sexual violent recidivism in adult male sexual offenders (Thornton et al., 2003). Information gathered via file review is used to score individuals on three scales: sexual recidivism (RM2000/S; 3 items), non-sexual violent recidivism (RM2000/V; 3 items), and combined recidivism (RM2000/C; 3 items). Total scores on the RM2000 scales are used to classify individuals into one of four risk categories (low, medium, high, very high), each of which has an estimated recidivism rate. Meta-analytic research has found all three RM2000 scales to be valid predictors of sexual, non-sexual violent, and general recidivism (Helmus, Babchishin, & Hanson, 2013).

**Static-99 and Static-99R.** The Static-99 and Static-99R are actuarial instruments designed to predict the likelihood of long-term sexual recidivism in adult sexual offenders (Hanson & Thornton, 1999; Helmus, Thornton, Hanson, & Babchishin, 2012). Information

gathered via file review is used to score individuals on 10 items, with the Static-99R having revised weights for the age at release item. Total scores on the Static-99 are used to classify individuals into one of four risk categories (low, moderate-low, moderate-high, high), each of which has an estimated recidivism rate. The Static-99 manual and forms are freely available via Public Service Canada ([www.ps.gc.ca](http://www.ps.gc.ca)), although the authors recommend training and ongoing quality assurance programs to ensure the measure is being used correctly. While training is recommended, the Static-99 can be scored by those without advanced professional degrees, as scoring is straightforward once the assessor is familiar with the scoring rules (Anderson & Hanson, 2010). Meta-analytic research and numerous primary studies have found the Static-99 and its revision to be valid and reliable predictors of sexual recidivism (Hanson & Morton-Bourgon, 2007; Singh, Fazel, Guerguieva, & Buchanan, 2013).

**Violence Risk Appraisal Guide (VRAG).** The VRAG is an actuarial instrument designed to predict the likelihood of violence in previously violent mentally disordered adult offenders (Quinsey, Harris, Rice, & Cormier, 2006). Information gathered via file review, third-party sources, and self-report is used to score 12 static items regarding childhood conduct, family background, previous criminal behavior, psychological issues, and index offense. These items were found to have the strongest statistical association with violent recidivism in a development sample of over 600 mentally disordered offenders (Harris et al., 1993). Assessors are expected to have training in the PCL-R, and should be able to demonstrate that they can score the VRAG reliably (Rice, Harris, & Hilton, 2010). Total scores on the VRAG are used to classify individuals into one of nine risk categories, each of which has an estimated recidivism rate. Replication studies have established the VRAG's ability to accurately predict violent recidivism in both men and women in a variety of settings as well as in child molesters, rapists, and non-violent offenders (Mental Health Centre Penetanguishene, 2009).

**Sex Offender Risk Appraisal Guide (SORAG).** The SORAG is an actuarial instrument designed to predict the likelihood of violent and sexual recidivism in convicted adult sexual offenders (Quinsey et al., 2006). Information gathered via file review, third-party sources, and self-report is used to score 14 items, 10 of which are identical to the VRAG and four additional items that are statistically related to sexual recidivism (e.g., phallometric deviance). Total scores on the SORAG are used to classify individuals into one of nine categories, each of which has an estimated recidivism rate. As with the VRAG, assessors are expected to have training in the PCL-R, and should be able to demonstrate that they can score the SORAG reliably (Rice, Harris, & Hilton, 2010). Studies conducted in a variety of study settings on a number of populations have found the SORAG to be a valid predictor of violent and sexual offending (Mental Health Centre Penetanguishene, 2009).

**Violence Risk Scale (VRS).** The VRS is an actuarial instrument designed to predict the likelihood of violence in adult forensic psychiatric patients (Wong & Gordon, 2009). Unique to the VRS is its incorporation of treatment-based change into actuarial risk estimates. Information gathered via file review and a semi-structured interview is used to score individuals on 26 items measuring two domains: static factors (6 items), and dynamic factors (20 items). Total scores on the VRS are used to determine an individual's risk level, with higher scores having a higher estimated recidivism rate. The VRS has been validated in samples of adult male offenders and forensic psychiatric inpatients (Burt, 2000, 2003; de Vries Robbé, Weenink, & de Vogel, 2006, Dolan & Fullam, 2007; Wilde & Wong, 2000; Wong & Gordon, 2006). In addition, meta-analytic research has found the static and dynamic factors on the VRS to be equally accurate in predicting violent recidivism (Yang, Wong, & Coid, 2010).

### 1.2.3 Structured Professional Judgment

Structured professional judgment (SPJ) risk assessment tools were developed to address the inflexibility of actuarial schemes. SPJ instruments are composed of risk and/or protective, static and/or dynamic factors that research or theory suggests are associated with the adverse event of interest. Total scores are used as an aide-memoire, guiding administrators in making a categorical risk judgment (e.g., Low, Moderate, or High) when combined with case-specific information gained through clinical experience with the client being evaluated. Hence, total scores are not to be used as statistical predictors of risk but rather as an important piece of a larger formulation process.

SPJ schemes seek to address the weaknesses of actuarial schemes. Thus, the key benefits of SPJ risk assessment tools include being more focused on individual clients than groups and the ability to take into consideration information not included in the item content of specific tools. The predictive validity of SPJ tools has been found to be non-significantly different than that of actuarial tools (Fazel, Singh, Doll, & Grann, 2012). In addition, practitioners generally perceive SPJ instruments to be more accurate and reliable than actuarial instruments and also of greater interest to mental health boards (Singh, 2013). The key drawbacks of SPJ risk assessment tools include a less objective evaluation process as well as the re-introduction of human decision-making biases into risk assessments. In addition, SPJ instruments are generally perceived as taking longer to administer than actuarial instruments (Singh, 2013).

Examples of commonly-used SPJ risk assessment schemes include the Brøset Violence Checklist (BVC; Almvik & Woods, 1999); Historical, Clinical, Risk Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997; Douglas, Hart, Webster, & Belfrage, 2013); Historisch, Klinisch, Toekomst-30 (HKT-30; Werkgroep, 2002); Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System (FOTRES; Urbaniok, 2007); Risk

Assessment, Management, and Audit System (RAMAS; O'Rourke, Hammond, Smith, & Davies, 1998); Structured Assessment of Protective Factors for Violence Risk (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbé, 2007, 2009, 2012); Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2006); Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009); Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997); and Violence Risk Screening-10 (V-RISK-10; Hartvig et al., 2007) .

**The Brøset Violence Checklist (BVC).** The BVC is an SPJ instrument designed to predict the likelihood of violence in adult psychiatric inpatients within the subsequent 24-hour period (Almvik & Woods, 1999). Information gathered via clinical observations is used to evaluate individuals on six items measuring two domains: patient characteristics (3 items) and patient behaviors (3 items). Users are not required to have any specific examiner qualifications, nor is training required to use the measure (Almvik, Woods, & Rasmussen, 2000). The item content on the BVC is based on the most frequent warning signals for violent inpatient incidents as identified by Linaker and Busch-Iversen (1995). Total scores on the BVC are used to classify individuals into one of three risk categories (small, moderate, very high), each of which has an estimated recidivism rate. In both longitudinal studies and randomized controlled trials, BVC assessments have been found to produce reliable and accurate assessments of short-term violence risk (Abderhalden et al., 2006; Almvik, 2008).

**Historical, Clinical, Risk Management-20 (HCR-20).** The HCR-20 is an SPJ instrument designed to predict the likelihood of violence in adult forensic and civil psychiatric patients (Webster, Douglas, Eaves, & Hart, 1997; Douglas, Hart, Webster, & Belfrage, 2013). Information gathered via file review and a semi-structured interview is used by a trained mental health professional to assess individuals on 20 items measuring three domains: historical factors (10 items), clinical factors (5 items), and risk management factors (5 items).

The HCR-20 has been validated in general and forensic psychiatric as well as correctional settings with adults across the lifespan (Douglas et al., 2013). Due to the dynamic nature of their item content, the clinical and risk management scales of the HCR-20 may be particularly useful in assisting professionals in making treatment decisions (Heilbrun, 2003). Recently, the authors released Version 3 of the measure (Douglas, Hart, Webster, & Belfrage, 2013) which retains the original 20 items, introduces some sub-items, and allows for assessor to rate on the relevance of items in addition to rating their presence or absence.

**Historisch, Klinisch, Toekomst-30 (HKT-30).** Modeled after the HCR-20, the HKT-30 was developed for use in the judicial context of Terbeschikkingstelling (TBS) in The Netherlands (Werkgroep, 2002). TBS is a provision allowing for mandatory psychiatric treatment for mentally disordered offenders following a prison sentence (van Marle, 2002). The HKT-30 is an SPJ instrument designed to predict the likelihood of violent recidivism in mentally disordered adult offenders. Information gathered via file review and a semi-structured interview is used at least two trained mental health professionals to assess individuals on 30 items measuring three domains: historical factors (11 items), clinical factors (13 items), and risk management factors (6 items). Item content on the HKT-30 takes into consideration regional risk factors such as difficulty with acculturation. Total scores on the HKT-30 are used to aid clinical judgment when classifying individuals into one of three risk categories (low, medium, high), with a recommended maximum duration of 12 months prior to reassessment. Systematic review evidence suggests that the HKT-30 produces accurate assessments of violent recidivism in both civil and forensic psychiatric patients (Singh, Serper, et al., 2011).

**Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System (FOTRES).** The FOTRES is an offense-specific, web-based SPJ instrument designed to assess and manage violent recidivism risk as well as to monitor treatment progress in adult



offenders (Urbaniok, 2007). Unique to this instrument is the prediction of reoffending with the same crime as the index offense. The FOTRES consists of approximately 650 items capturing four domains: structural risk of recidivism, suggestibility, therapy evaluation, and treatment prognosis. These four domains establish a Risk Needs Level as well as a Risk Management Level: The Risk-Needs Level is only assessed at the time of the index offense and not only estimates recidivism risk but also determines treatment outcome. The Risk Management Level is assessed on multiple occasions to measure treatment progress and changes to recidivism risk associated with situational factors. As its item content focuses on tailoring risk management plans, the FOTRES may be useful for identifying treatment targets as well as for maximizing responsivity (Studer, Aylwin, Stribney, & Reddon, 2011). Primary research findings on the predictive validity and inter-rater reliability of the FOTRES has been mixed (Rossegger et al., 2011; Keller et al., 2011).

**Risk Assessment, Management, and Audit System (RAMAS).** The RAMAS is an SPJ instrument designed to predict the likelihood of violence in mental health, criminal justice, and social service users (O'Rourke, Hammond, Smith, & Davies, 1998). Information gathered using a multidisciplinary and interagency approach is used to evaluate individuals on 66 items measuring four domains: risk to self, risk to others, mental health risk, and vulnerability. The tool is intended for collaborative use by professionals in hospital, institutional, community, and research settings to achieve the twin objectives of providing better protection to the public whilst improving care planning, treatment, and management of people posing a risk to themselves or others. Multi-disciplinary training is structured across three levels, including a half-day practitioner workshop (Level 1), training as a RAMAS risk and care trained user (Level 2), and a License to Train (Level 3). Users must complete Levels 1 and 2 in order to use the RAMAS system. According to the instrument's developers, RAMAS is also an acronym for the goal of the system: Realistic, Achievable, Measurable, Appropriate

Strategies for Health and Well-being. Research on the RAMAS Risk Assessment Checklist has shown that the various item scales demonstrate internal consistency estimates of reliability that are within the recommended range for the Checklist to be considered reliable. A high degree of consistency between raters has also been demonstrated, although correlations between RAMAS scales and clinical judgment are noted to be either not statistically significant, or not large. When examined as separate scales, (i.e., Dangerousness, Mental Instability, and Self Harm), the psychometric properties of the scales are excellent, and allow for discriminating risk assessment profiles to be drawn for individual patients (Hammond & O'Rourke, 2000).

**Structured Assessment of Protective Factors for Violence Risk (SAPROF).** The SAPROF is an SPJ instrument that was developed to supplement the HCR-20 and similar instruments by adding protective factors for violence into the case formulations of adult violent and sexual offenders (de Vogel, de Ruiters, Bouman, & de Vries Robbé, 2007, 2009, 2012). Information gathered via file review and a semi-structured interview is used to score individuals on 17 items measuring three domains: internal factors (5 items), motivational factors (7 items), and external factors (5 items). The SAPROF has been found to be a reliable and valid supplement to both the HCR-20 Version 2 (de Vries Robbé, de Vogel, Douglas, & Nijman, in press) as well as the HCR-20 Version 3 (de Vries Robbé, de Vogel, Koster, & Bogaerts, 2014), with more accurate assessments being produced for violent rather than sexual recidivism. Research examining the SAPROF has demonstrated excellent levels of inter-rater reliability and good levels of predictive validity for violent recidivism (de Vries Robbé & de Vogel, 2013; de Vries Robbé, de Vogel, & de Spa, 2011; de Vries Robbé, de Vogel, Koster, & Bogaerts, 2015).

**Structured Assessment of Violence Risk in Youth (SAVRY).** The SAVRY is an SPJ instrument designed to assess the likelihood of violence in male and female adolescents

aged 12 to 18 years (Borum, Bartel, & Forth, 2006). Information gathered via file review and a semi-structured interview is used to evaluate 30 factors measuring four domains: historical factors (10 items), social/contextual factors (6 items), individual/clinical factors (8 items), and protective factors (6 items). The SAVRY allows raters to designate risk factors as “critical”, if the rater judges them to be causal in the given case. Total scores on the SAVRY are used to aid clinical judgment when classifying individuals into one of three risk categories (low, moderate, high). As it includes dynamic risk and protective factors, the SAVRY may be a useful aid to intervention and management planning in addition to risk assessment (Borum, Lodewijks, Bartel, & Forth, 2010). Professionals in a variety of disciplines who conduct violence risk assessments or prepare risk management plans are qualified to use the SAVRY, and there are no specific training requirements for using the instrument (Borum et al., 2010). Research on the SAVRY has demonstrated good to excellent levels of inter-rater reliability and good levels of internal consistency (Catchpole & Gretton, 2003; Dolan & Rennie, 2008; Lodewijks, Doreleijers, de Ruiter & Borum, 2008; McEachran, 2001; Meyers & Schmidt, 2008; Viljoen et al., 2008). Meta-analytic research has found the SAVRY to produce rates of predictive validity that exceed those of risk assessment tools developed for adult populations (Singh, Grann, & Fazel, 2011).

**Short-Term Assessment of Risk and Treatability (START).** The START is an SPJ instrument designed to predict the likelihood of violence, suicide, unauthorized leave, victimization, self-harm, substance abuse, and self-neglect in the short-term (weeks to months; Webster, Martin, Brink, Nicholls, & Desmarais, 2009). The START is intended for use by experienced mental health and correctional professionals to assess a diverse set of adult populations and settings (e.g., civil and forensic psychiatric patients, correctional inmates, probationers). Information gathered via file review and a semi-structured interview is used to concurrently score individuals on 20 items measuring two domains: strengths and

vulnerabilities. The START allows raters to designate vulnerabilities as “critical”, if the rater judges a vulnerability to be a causal risk factor. The START allows raters to designate strengths as “key”, if the rater judges a strength to be a causal protective factor. Total scores on the START are used to aid clinical judgment when classifying individuals into one of three risk categories (low, moderate, high) for the instrument’s seven outcomes of interest. Webster and colleagues (2009) recommend that the duration of time between assessments using this dynamic measure not exceed three months. A recent meta-analysis concluded that START assessments of violence risk produce excellent levels of inter-rater reliability and fair to excellent levels of predictive validity (O’Shea & Dickens, 2014). However, there is currently a lack of research on the accuracy of START assessments when used to predict the likelihood of future suicide, substance abuse, and unauthorized leave.

**Sexual Violence Risk-20 (SVR-20).** The SVR-20 is an SPJ instrument designed to assess the likelihood of sexual violence in adult male sex offenders (Boer, Hart, Kropp, & Webster, 1997). The instrument’s authors have stated that the SVR-20 may be used with careful consideration with women as well as with older male adolescents (16 or 17 years old (Hart & Boer, 2010). Information gathered via file review and a semi-structured interview is used to evaluate individuals on 20 items measuring three domains: psychosocial adjustment (11 items), sexual offenses (7 items), and future plans (3 items). Total scores on the SVR-20 are used to aid clinical judgment when classifying individuals into one of three risk categories (low, moderate, high). According to the SVR-20 manual, users of the scheme should meet two general requirements: (1) basic familiarity with professional and scientific literature on sexual violence, including its causes and management; and (2) training and experience in individual assessment, including interviewing and file review (Hart & Boer, 2010). A recent narrative review concluded that, in light of the findings of replication studies, the SVR-20 should be considered a valid and reliable predictor of violent recidivism in sexual offenders

(Rettenberger, Hucker, Boer, & Eher, 2009). Research on the SVR-20 has established good to excellent levels of inter-rater reliability (de Vogel, de Ruiter, van Beek & Mead, 2004; Rettenberger & Eher, 2007; Watt & Jackson, 2008; Zanatta, 2005), strong concurrent validity (Dietiker, Dittmann, & Graf, 2007; Rettenberger & Eher, 2007; Zanatta, 2005), and variable predictive validity (Barbaree et al., 2008; Craig, Browne, Beech, & Stringer, 2006; Dempster, 1998; de Vogel et al., 2004; Sjöstedt & Långström, 2003; etc.).

**Violence Risk Screening-10 (V-RISK-10).** The V-RISK-10 is an SPJ instrument designed to predict the likelihood of violence in civil psychiatric patients such that a more comprehensive risk assessment may be conducted or immediate risk management plans can be put in place (Hartvig et al., 2007). The measure is intended for use by psychologists, psychiatrists, and non-psychiatrist medical doctors. Information gathered via file review and clinical observation is used to score individuals on 10 items measuring three domains: historical, clinical, and risk management items. Total scores on the V-RISK-10 are used to aid clinical judgment when classifying individuals into one of three risk categories (low, moderate, or high). Prospective studies examining the V-RISK-10 have found the instrument to be as reliable and valid as more comprehensive risk assessment instruments (Bjørkly, Hartvig, Roaldset, & Singh, 2014), with particular usefulness in identifying low risk patients (Centre for Research and Education in Forensic Psychiatry, 2008; Hartvig et al., 2011).

### 1.3 Integrating Risk Assessment Research into Practice

Over the past 30 years, more than 400 forensic risk assessment tools have been developed for the purposes of predicting the likelihood of future violence, sex offending, and general recidivism (Singh, 2013). In accordance with a recent amicus curiae brief from the American Psychological Association (Gilfoyle et al., 2011), it is recommended that such structured instruments be routinely used by mental health and criminal justice professionals and that judges and lawyers seek out evaluators who use such instruments rather than

unstructured clinical judgments. This said, risk assessment tools should not be the sole determinants of decisions concerning civil liberties, especially when the base rate of the outcome of interest is particularly low (McSherry & Keyzer, 2009).

Though currently used in over 40 countries for prediction, management, and monitoring purposes (Singh, 2013), no single risk assessment tool has emerged as being more accurate than others (Yang, Wong, & Coid, 2010). To decide which tool to use, meta-analytic research suggests focusing on the intended population and outcome for which a tool was designed, and then trying to find a “best fit” with the population and outcome of interest (Singh, Grann, & Fazel, 2011). The more deviations from a tool manual (e.g., item omissions, changes in scoring procedures), the weaker the tool’s performance – this extends to using a “clinical override” on estimates established by actuarial risk assessment tools (Quinsey et al., 2006).

As new risk assessment tools have recently been developed for more specific populations – for example, intellectually disabled offenders (Lofthouse, Lindsay, Totsika, Hasting, & Roberts, 2014) – to assess the likelihood of more specific outcomes – for example, spousal assault (Kropp, Hart, & Belfrage, 2010) and suicide (Stegg et al., 2012) – there has been a renewed focus on moving beyond static risk factors and moving towards incorporating more

#### 1.4 The Present Thesis

Risk assessment protocols have been implemented in mental health and criminal justice settings around the globe to prioritize risk reduction strategies for those most at need. Helping to allocate scarce resources more effectively and efficiently while protecting our communities, risk assessment has come to be a cornerstone of forensic practice in many jurisdictions. The present thesis investigates the use, perceived utility, and research quality on forensic risk assessment tools. First, a systematic review of qualitative surveys on forensic

risk assessment tools is conducted to explore the international use and perceived utility of such instruments (Article 1). The systematic review concludes that there is a lack of methodologically-rigorous qualitative research on forensic risk assessment tool use and perceived utility across continents and professional disciplines, necessitating a new, large-scale study (Article 2). Given the lack of methodologically-rigorous qualitative research on forensic risk assessment tools, a second systematic review is conducted to explore the methodological rigor of the quantitative research literature concerning the predictive validity of these instruments (Article 3). The systematic review concludes that the field could benefit from the development of standardized guidelines promoting the transparent and consistent reporting of methodology and results in peer-reviewed publications on violence risk assessment tools (Article 4). These standardized guidelines are then applied to the sexual offender risk assessment tool research literature to examine its transparency and consistency of reporting of methodology and results (Article 5).

**Table 1:** Characteristics of Commonly-Used Risk Assessment Tools

Tool	Characteristics				
	Items	Approach	Intended Population(s)	Intended Outcome(s)	Time (Minutes)
BVC	6	Actuarial	Psychiatric Acute Inpatients	Violence	<5
COVR	Variable	Actuarial	Acute Psychiatric Inpatients	Violence	10-15
FOTRES	Variable	SPJ	General Offenders	Variable	60
HCR-20	20	SPJ	Civil/Forensic Psychiatric Patients	Violence	Unknown
HKT-30	30	SPJ	Mentally Disordered Violent Offenders	Violent Recidivism	Unknown
LS/CMI	43	Actuarial	General Offenders	General Recidivism	20-30 per form (5 forms)
LSI-R	54	Actuarial	General Offenders	General Recidivism	30-45
PCL-R	20	N/A	Forensic Populations	N/A	90-120 (Interview); 60 (File Review)
PCL:SV	12	N/A	Psychiatric Acute Inpatients	N/A	45 (Interview); 30 (File Review)
RAMAS	66	SPJ	Mental Health Service Users	Risk to Others; Risk to Self; Victimization	Unknown
RM2000	7	Actuarial	Sexual Offenders	Sexual, Non-sexual Violent, or Combined Recidivism	Unknown
SAPROF	17	SPJ	Violent and Sexual Offenders	Violence	Unknown
SAVRY	30	SPJ	Adolescents	Violence	Unknown
SORAG	14	Actuarial	Sexual Offenders	Violent Recidivism	Unknown
Static-99	10	Actuarial	Sexual Offenders	Sexual Recidivism	15
Static-99R	10	Actuarial	Sexual Offenders	Sexual Recidivism	15
START	20	SPJ	Civil/Forensic Psychiatric Patients	Violence; Suicide; Self-Harm; Unauthorized Leave; Victimization; Substance Abuse; Self-Neglect	22-25
SVR-20	20	SPJ	Sexual Offenders	Violent Recidivism	Unknown
V-RISK-10	10	SPJ	Non-forensic Psychiatric Patients	Violence	Unknown
VRAG	12	Actuarial	Mentally Disordered Violent Offenders	Violent Recidivism	Unknown
VRS	26	Actuarial	Forensic Psychiatric Inpatients	Violent Recidivism	Unknown



**Table 2:** Types of Items Included in Commonly-Used Risk Assessment Tools

Tool	Types of Items			
	Risk	Protective	Static	Dynamic
BVC	X			X
COVR	X	X	X	X
FOTRES	X		X	X
HCR-20	X		X	X
HKT-30	X		X	X
LS/CMI	X		X	X
LSI-R	X		X	X
PCL-R			X	X
PCL:SV			X	X
RAMAS	X		X	X
RM2000	X		X	
SAPROF		X	X	X
SAVRY	X	X	X	X
SORAG	X		X	
Static-99	X		X	
Static-99R	X		X	
START	X	X		X
SVR-20	X		X	X
V-RISK-10	X		X	X
VRAG	X		X	X
VRS	X		X	X

**Table 3:** Item Content Included in Commonly-Used Risk Assessment Tools

Tool	Item Content Domain									
	Attitudes	Associates/ Peers	History of Antisocial Behavior	Personality Problems	Relationships	Work/ School	Recreation/ Leisure Activities	Substance Use	Mental Health	Housing Status
BVC			X						X	
COVR	X	X	X	X	X	X		X	X	X
FOTRES	X	X	X	X	X	X		X	X	X
HCR-20	X	X	X	X	X	X		X	X	
HKT-30	X	X	X	X	X	X	X	X	X	X
LS/CMI	X	X	X	X	X	X	X	X	X	
LSI-R	X	X	X	X	X	X	X	X	X	X
PCL-R	X	X	X	X	X				X	
PCL:SV	X	X	X	X	X				X	
RAMAS	X		X	X	X			X	X	
RM2000			X		X					
SAPROF	X	X			X	X	X		X	X
SAVRY	X	X	X		X	X	X	X	X	
SORAG			X	X	X	X		X	X	
Static-99			X		X					
Static-99R			X		X					
START	X	X			X	X	X	X	X	
SVR-20	X		X	X	X	X		X	X	
V-RISK-10			X	X				X	X	X
VRAG			X	X	X	X		X	X	
VRS	X	X	X	X	X	X		X	X	

## 2 RESEARCH ARTICLES

---

2.1 Hurducas, C., **Singh, J. P.**, de Ruiter, C., & Petrila, J. (2014). Violence risk assessment tools: A systematic review of clinical surveys. *International Journal of Forensic Mental Health, 13*, 181-192.

### 2.1.1 Abstract

The present study is a systematic review exploring the methodological quality and consistency of findings for surveys on the use of violence risk assessment tools. A systematic search was conducted to identify surveys of violence risk assessment tool use published between January 1, 2000 and January 1, 2013 using PsycINFO, MEDLINE, and EBSCO Criminal Justice Abstracts. Characteristics of survey administration and more findings were extracted, and a checklist of 26 reporting quality markers in survey research was used for coding. Nine surveys were identified, fulfilling on average approximately half of the quality markers ( $M = 15.5$ ,  $SD = 1.6$ ). An average of 104 respondents ( $SD = 93$ ) participated, with a range of 10 to 300 respondents. Most surveys examined the practices of psychologists in the United Kingdom or the United States. The Psychopathy Checklist-Revised and the Historical, Clinical, Risk Management-20 were the most commonly used instruments by practitioners. No surveys investigated differences in assessment practices across professional disciplines or continents, and none examined the use or perceived usefulness of structured instruments in risk management or risk monitoring. There continues to be a need for transparent, high quality clinical surveys on the use and perceived utility of violence risk assessment tools in the forensic mental health field. Given the growing cross-jurisdictional use of risk assessment tools, comparisons of international practice are particularly important.

### 2.1.2 Introduction

The prevention of violence has generated considerable clinical and research interest. After seminal research found that unstructured judgments of risk were no more valid than chance several decades ago (Monahan, 1981; Steadman & Cocozza, 1974; Thornberry & Jacoby, 1979), a number of tools have been developed to add structure to the process of assessing future violence risk and, thus, increase its predictive accuracy. According to a recent systematic review (Singh, Serper, Reinharth, & Fazel, 2011), over 150 risk assessment tools have specifically been developed to assess the risk of violence. These instruments include schemes such as the Historical, Clinical, Risk Management-20 (HCR- 20; Webster, Douglas, Eaves, & Hart, 1997), the Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993), and the Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997). Such instruments are now used on multiple continents including North America (Bloom, Webster, Hucker, & De Freitas, 2005), South America (Folino & Castillo, 2006), Europe (Risk Management Authority, 2007), Africa (Roffey & Kaliski, 2012), Asia (Chu, Daffern, Thomas, & Lim, 2012), and Australia (Ogloff & Davis, 2005).

Numerous studies have investigated the psychometric properties of risk assessment tools in the context of research, but comparatively few have explored how these instruments are actually used in practice (Elbogen, 2002). Such information is important, as evidence suggests that using risk assessment tools with their intended population to predict their intended outcome of interest can maximize reliability and predictive validity (Harris & Rice, 2007; Singh, Grann, & Fazel, 2011). Hence, examining what tools are used in the field and how is of arguably greater practical importance than their utility in controlled research contexts. One approach to examining the application of violence risk assessment tools in practice is through survey methodology.

## **Survey Research**

A survey is a systematic data collection tool used to gather information from a representative sample of a population which can be generalized to that entire population (Groves et al., 2009). There are four types of surveys generally used: face-to-face interviews, telephone questionnaires, postal mail questionnaires, and Web-based questionnaires. Face-to-face interviews are the most direct and intrusive form of surveying, but they have also been found to yield the highest response rates (Hox & De Leeuw, 1994; Krysan, Schuman, Scott, & Beatty, 1994). However, such interviews are limited in their utility by a high likelihood of researcher reactivity effects, time constraints, and safety issues for interviewers (Babbie, 2012). Telephone questionnaires address this by being less time-consuming and avoiding physical proximity between interviewer and respondent, though they can be more expensive and limit samples to persons who own and actively answer their phone (Holbrook, Green, & Krosnick, 2003). Postal mail questionnaires are relatively inexpensive and allow respondents to complete surveys at their convenience, taking as much time as needed. However, such surveys have been found to have lower response rates compared to telephone and face-to-face methods (Cobanoglu, Warde, & Moreo, 2001). Finally, Web-based questionnaires reduce the time and costs associated with surveying, with the added benefit of avoiding the often error-prone and tedious task of data entry (Medin, Roy, & Ann, 1999). However, electronic approaches to surveying suffer from coverage bias, as they can only recruit individuals who have access to the internet (Crawford, Couper, & Lamias, 2001). The extent of this bias will depend on the age, socioeconomic status, and geographic location of the population of interest (Kaplowitz, Hadlock, & Levine, 2004).

## **Present Study**

Though a recent metareview identified a number of systematic reviews that have been conducted to investigate the psychometric properties of violence risk assessment tools (Singh

& Fazel, 2010), none have reviewed the survey literature on such instruments. Hence, the aim of the present study was to examine the transparency and consistency of published surveys concerning violence risk assessment tool use in practice. Specifically, we wished to explore the quality of these surveys and to identify gaps in knowledge that future survey research could address.

### 2.1.3 Method

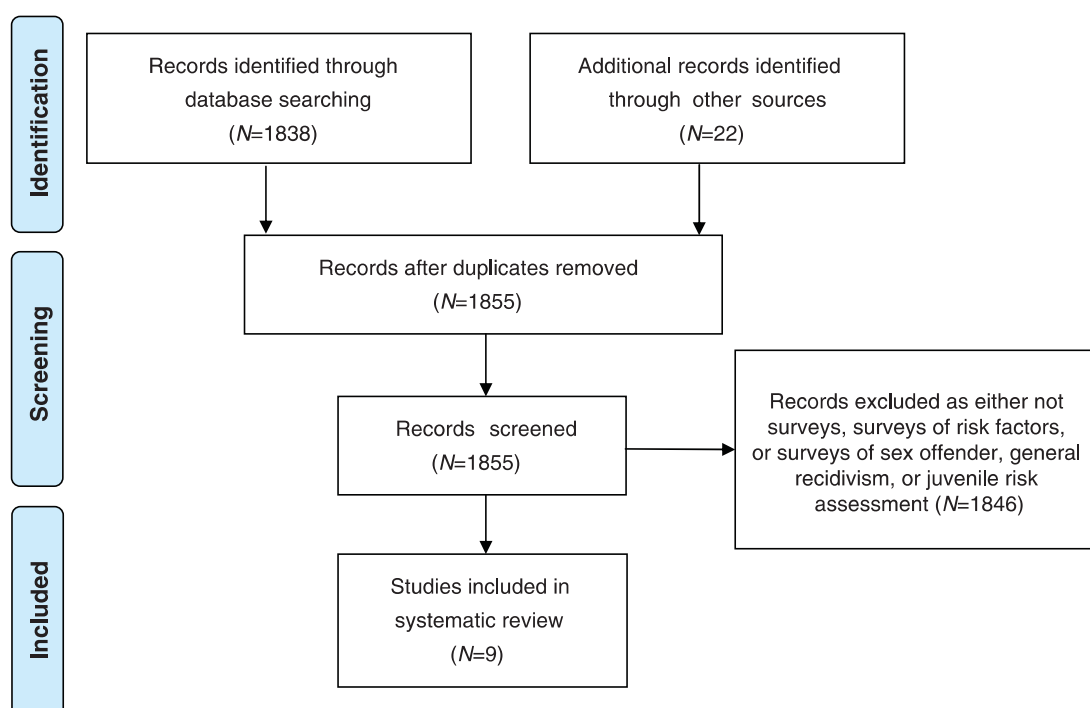
#### **Review Protocol**

For a consistent and transparent reporting of results, the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) Statement was followed (Moher, Liberati, Tetzlaff, & Altman, 2009). This 26-item checklist ensures accurate reporting of review methodology and findings.

#### **Systematic Search**

A systematic literature search was conducted using PsycINFO, MEDLINE, and EBSCO Criminal Justice Abstracts to identify English-language surveys of violence risk assessment tool use and perceived utility. Only studies published between January 1, 2000 and January 1, 2013 were included as we sought to explore contemporary practice. Combinations of the following Boolean keywords were used: violen\*, risk, assessment, prediction, and survey. Additional surveys were identified through reference sections, annotated bibliographies, and correspondence with risk assessment researchers. As a quality control measure, only surveys published in peer-reviewed journals were considered for inclusion. Surveys concerning alternative areas of forensic risk assessment such as sex offender risk assessment (e.g., Jackson & Hess, 2007) or general recidivism risk assessment (e.g., Taxman, Cropsey, Young, & Wexler, 2007) were excluded, as were surveys on juvenile risk assessment (e.g., Shook & Sarri, 2007), communication of violence risk (Heilbrun et al., 2004), and risk factors for violence (e.g., Elbogen, Mercado, Scalora, & Tomkins, 2002). The

initial search identified 1,855 records (Figure 2). When screened for eligibility, nine studies were found to meet inclusion and exclusion criteria (Archer, Buffington-Vollum, Stredny, Handel, 2006; Bengtson & Pedersen, 2008; Green, Carroll, & Brett, 2010; Hawley, Gale, Sivakumaran, & Littlechild, 2010; Higgins, Watts, Bindman, Slade, & Thornicroft, 2005; Khuroya, Weaver, & Maden, 2009; Lally, 2003; Tolman & Mullendore, 2003; Viljoen, McLachlan, & Vincent, 2010).



**Figure 1:** Systematic search for surveys investigating violence risk assessment tool use and perceived utility.

## Data Extraction

The objectives of the present study were: to assess the quality of the surveys included in the review, to explore how each was administered, and to analyse their findings. Three different data extraction methods were used to achieve these objectives: a quality checklist (Bennet et al., 2011), an administration characteristics coding sheet, and a findings coding sheet.

A 26-item checklist was used to assess the reporting quality of each survey (Table 4). This checklist was developed by Bennett and colleagues (2011) who reviewed recent surveying guidelines (e.g., Burns et al., 2008; Draugalis, Coons, & Plaza, 2008; Kelley, Clark, Brown, & Sitzia, 2003) and identified key reporting domains, all considered equally important to the checklist authors.

A supplementary coding sheet was developed by the authors to assist in the extraction of characteristics on survey administration findings. The eight extracted characteristics included: 1. Language(s) in which the survey was administered; 2. Number of professional disciplines sampled; 3. Method of survey administration; 4. Number of organizations to which the survey was disseminated; 5. Number of reminders sent to encourage participation; 6. Incentives offered to respondents; 7. Inclusion of survey items concerning the use of tools in risk assessment, risk management, and/or risk monitoring; 8. Reported survey items concerning the perceived usefulness of tools in risk assessment, management, and/or monitoring. To explore key survey findings, the following 10 respondent characteristics were extracted using a third coding sheet: 1. Number of respondents; 2. Response rate; 3. Number of countries represented by respondents; 4. Mean age of respondents; 5. Percentage of respondents that were male; 6. Number of risk assessments conducted by respondents over lifetime; 7. Percentage of risk assessments conducted over life-time using a structured tool; 8. Number of risk assessments conducted by respondents over past 12 months; 9. Percentage of risk assessments conducted over past 12 months using a structured tool; 10. The three structured tools most commonly used by respondents



As a measure of quality control for the data extraction, five (55.5%) of the included studies were randomly selected and coded by the second author. This investigator was provided with the quality checklist, the standardized coding sheets, and the five study manuscripts. Using Cohen's (1960) kappa, a perfect level of interrater agreement was established ( $k = 1.00$ ; Landis & Koch, 1977).

### **Procedure**

Descriptive analyses were conducted to examine trends in the distribution of quality markers and differences in survey administration and findings. In addition, a narrative overview was prepared for each survey to summarize findings related to violence risk assessment.

**Table 4:** Methodological Quality Markers in Nine Clinical Surveys of Use of Structured Violence Risk Assessment Instruments

Methodological Quality Marker	Archer et al. (2006)	Bengtson & Pedersen (2008)	Green et al. (2010)	Hawley et al. (2010)	Higgins et al. (2005)	Khiroya et al. (2009)	Lally (2003)	Tolman & Mullendore (2003)	Viljoen et al. (2010)
<b>Background</b>									
Justification of research method	–	–	–	–	–	–	–	–	–
Background literature review	•	•	•	•	•	•	•	•	•
Explicit research questions	–	•	–	•	–	–	•	•	•
Clear study objectives	•	–	•	•	•	•	–	•	•
<b>Methods</b>									
Description of methods used for data analysis	•	•	•	•	•	•	•	–	•
Method of questionnaire administration	•	•	•	•	•	•	•	•	•
Number and types of contact	–	•	–	–	•	•	•	•	•
<b>Sample selection</b>									
Sample size calculation	•	•	•	–	•	•	•	•	•
Representativeness	–	•	–	–	•	–	•	–	–
Method of sample selection	•	–	•	•	•	•	•	•	•
<b>Research tool</b>									
Description of the research tool	•	•	–	•	–	•	•	•	•
Instrument pretesting	–	–	–	•	–	–	–	–	–
Instrument reliability and validity	–	–	–	–	–	–	–	–	–
<b>Results</b>									
Results of research presented	•	•	•	•	•	•	•	•	•
Results address objectives	•	–	•	•	•	•	–	•	•
Generalizability	•	•	–	•	–	–	–	–	–
<b>Response rates</b>									
Response rate stated	•	•	–	•	•	•	•	•	•
How response rate was calculated	•	•	–	–	•	•	•	•	•
Discussion of nonresponse bias	–	–	–	•	–	•	–	–	–
<b>Interpretation and discussion</b>									
Interpret and discuss findings	•	•	•	•	•	•	•	•	•
Conclusion	•	•	•	•	–	–	–	•	•
Recommendations	–	•	•	•	•	•	–	•	•
Limitations	•	•	–	•	•	•	•	–	•
<b>Ethics and disclosure</b>									
Consent	–	–	–	–	–	–	–	–	–
Sponsorship	–	–	•	–	•	–	–	–	•
Research ethics approval	–	–	–	•	–	–	–	–	–

*Note.* • = characteristic present in survey; – = characteristic absent from survey.

## 2.1.4 Results

### Survey Reporting Quality

The nine included studies were screened for markers of reporting quality detailed in Table 4. The average survey met just over half of the criteria ( $M = 15.5$ ,  $SD = 1.6$ , range = 12–18). Of the sections constituting the checklist, the Methods section was most completely reported across studies ( $M = 2.5$  of 3 criteria met,  $SD = 0.5$ , range = 2–3) followed by the Interpretation and Discussion section ( $M = 3.2$  of 4,  $SD = 0.6$ , range = 2–4), the Sample Selection and the Results section (both  $M = 2.1$  of 3,  $SD = 0.6$ , range = 1–3), the Response Rate section, ( $M = 1.8$  of 3,  $SD = 0.7$ , range 0–3), and the Background section ( $M = 2.4$  of 4,

SD = 0.5, range = 2–3). The Research Tool (M = 0.8 of 3, SD = 0.6, range = 0–2) and Ethics and Disclosure (M = 0.4 of 3, SD = 0.5, range = 0–1) sections had the fewest criteria met.

The most commonly reported items across sections were: the method of questionnaire administration, background literature review, results of survey research, and interpretation and discussion of findings (N surveys meeting criteria = 9, 100%). None of the included studies described the procedure through which informed consent was obtained. In addition, no studies investigated the test-retest reliability or face validity of the survey that was administered, although one study did describe a pilot phase to ensure that all questions were relevant and clear in their phrasing (Hawley et al., 2010). The most complete survey satisfied 18 (69.2%) of the reporting quality markers (Viljoen et al., 2010).

### **Characteristics of Survey Administration**

Eight characteristics concerning surveying administration were extracted from each of the included studies (Table 5). The number of professional disciplines sampled by each study varied from one to five, with psychologists being the most common profession (N = 5; 62.5%). Web-based questionnaires were administered in three (33.3%) studies, with the remaining six (66.6%) studies using a paper-and-pencil format. An average of 1.8 sources (SD = 1.3; range = 1– 5)—most commonly online directories or ListServes of professional organizations—were used by researchers to disseminate surveys. An average of 1.6 reminders (SD = 0.5; range = 1–2) were sent to respondents to encourage participation. Only three (33.3%) of the included studies measured perceived usefulness of risk assessment tools (Hawley et al., 2010; Khiroya et al., 2009; Viljoen et al., 2010), and only one (12.5%) surveyed the frequency of risk assessment tool use in risk management (Bengtson & Pedersen, 2008). None of the studies reported whether they offered incentives to participants or not, or ratings of the perceived utility of risk assessment tools for risk management or risk monitoring.

### **Characteristics of Survey Findings**

Ten characteristics concerning the participants who completed the surveys and their responses were extracted. The average number of respondents was 104 (SD = 93, range = 10–300). The average response rate was 55.8% (SD = 13.8%, range = 35–83%). Seven of the nine surveys targeted specific countries, most commonly the United Kingdom (N = 3; 33.3%) and the United States (N = 2; 22.2%). Although the gender composition and average age of samples were only reported in four surveys, there was a trend towards middle-aged men comprising the majority of respondents. None of the surveys reported the average number of risk assessments conducted by respondents over their lifetime; however, one study reported a median of 55 risk assessments using structured tools conducted in the previous 12 months (Green et al., 2010). The Psychopathy Checklist-Revised (PCL-R; Hare, 2003) and its screening version (PCL:SV; Hart, Cox, & Hare, 1995) and the HCR- 20 were the most commonly used tools when frequency of tool use was described.

### **Descriptions of Surveys**

*Archer et al. (2006)*

Archer and colleagues used a Web-based survey to explore which psychological tests are used in forensic evaluations by 152 doctoral-level members of the American Psychology-Law Society and diplomates of the American Board of Forensic Psychology. The survey explored the use of 10 categories of instruments: measures of psychopathy and risk assessment, sexual offender scales, clinical scales, child-related forensic instruments, malingering tests, measures of competency and insanity, multiscale personality inventories, unstructured personality tests, neuropsychological assessment, and cognitive/school achievement tools. In addition to whether they used each category of instrument, respondents also indicated how often they used each instrument. The researchers found that the most commonly and frequently used risk assessment tools for evaluating violence were the PCL-R,

the HCR-20, and the VRAG. The authors concluded that the use of specialized instruments such as the HCR-20 and the VRAG had increased compared to previous surveys conducted within the field of violence risk assessment.

*Bengtson & Pedersen (2008)*

Bengtson and Pedersen conducted a survey of 41 psychologists and psychiatrists working in Danish forensic psychiatric units to examine the use of violence risk assessment tools in mental health evaluations. Respondents were found to be most familiar with the PCL-R, PCL:SV, HCR-20, SVR-20, and VRAG, whereas the PCL-R, and PCL:SV were the most commonly used instruments specifically for violence risk assessment. The most commonly reported reason for using risk assessment instruments was the desire for evidence-based practice, and the most commonly reported reasons for not using them were insufficient training and the preference for unstructured clinical judgment. Amongst those clinicians who reported using a risk assessment instrument, an average of 3.8 instruments (SD = 2.6, range = 2–8) were used in this process over the course of their careers.

*Green et al. (2010)*

Green and colleagues surveyed the use of the HCR-20 by community forensic mental health services in Australia. Clinical teams within these units included psychologists, psychiatrists, nurses, social workers, and registrars. They sent a mail questionnaire to all 12 such units in Australia to obtain comparative data on the use of the HCR-20. The researchers found that approximately 50 HCR-20 assessments had been completed by each team over the past 12 months. There was considerable variation across services in the number of assessments performed, the amount of time allocated for gathering information and report writing, and the type of training offered in administering the tool. The researchers concluded that SPJ instruments such as the HCR-20 are time-consuming in terms of administration and

interpretation and that training is highly important to ensure time-effectiveness, reliability, and rating consistency when implementing such measures.

*Hawley et al. (2010)*

Hawley and colleagues conducted a survey of 300 practitioners (self-classified as doctors, nurses, or other health professionals) working in the Hertfordshire National Health Service Trust in the United Kingdom. The practitioners were asked to describe the amount of time taken to complete the locally-agreed risk assessment proforma (RAP; a standardized form) and their attitudes about the usefulness of such structured measures. The researchers found that the time allocated to complete RAPs is under 2% of the average working day, but there was considerable variability in estimates. Three-fourths of respondents reported that RAPs were useful in at least half of cases for which they were completed. The survey found that doctors allocated less time to completing RAPs and viewed them as less useful than nurses. However, differences in specific instrument use and perceived usefulness across disciplines were not examined.

*Higgins et al. (2005)*

Higgins and colleagues conducted a survey of consultants working in 66 randomly selected mental health trusts across England to establish current violence risk assessment practices in general adult psychiatry. The researchers developed a brief semi-structured questionnaire to explore whether structured or unstructured methods were used in each trust. They also requested copies of each trust's risk assessment protocols, when available. The researchers found that the majority of the sampled trusts had developed their own standardized risk assessment protocols rather than adopting available validated schemes. Approximately half of the trusts offered routine training in the use of their self-developed protocols, although attendance at these trainings was not as high as expected. A content analysis of the protocols sent to the researchers revealed wide variation in included risk

factors as well as a lack of standardization in how final scores were interpreted or categorical estimates assigned. Approximately half of the forms incorporated a plan for managing identified risk.

*Khiroya et al. (2009)*

Khiroya and colleagues conducted a survey of the clinical service directors at 47 adult medium secure forensic units across the United Kingdom to explore the use of violence risk assessment instruments. Respondents reported that a variety of violence risk assessment tools had been implemented and that most units routinely used more than one. The most commonly used instruments were the PCL-R and the HCR-20. The researchers conducted unstructured follow-up interviews with a subset of respondents to ask about perceived tool utility. Respondents reported using structured risk assessment instruments as part of a wider battery of structured assessment tools. The most frequently offered reasons for adopting a specific instrument were research evidence and encouragement by local trusts. The Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009) was rarely used, with only one unit reporting using it frequently, but the instrument was judged to be the most useful according to interviewed directors. For measuring perceived utility, a global usefulness rating was used, unrelated to any specific task (i.e., risk assessment, risk management, or risk monitoring).

*Lally (2003)*

Lally surveyed the general acceptability of structured assessment instruments in forensic evaluations as rated by 64 psychologists who were diplomates of the American Board of Forensic Psychology. In the United States, such acceptability is a criterion to establish whether expert court testimony aided by an instrument is scientifically valid under different legal standards, such as Daubert (Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993) or Frye (Frye vs. United States, 1923). Thus, respondents were asked to rate the

acceptability of the instruments used to address six categories of psycholegal issues: violence risk, sexual violence risk, mental state at the time of the offense, competency to stand trial, competency to waive Miranda rights, and malingering. The researchers found that the PCL-R, the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, Kraemmer, 1989), and the PCL:SV were the three instruments perceived to be most acceptable in the violence risk assessment process. They also concluded that although an instrument may be recommended by a majority of forensic psychologists, this alone does not mean that it is commonly used in practice. For example, respondents rarely endorsed as recommended a number of instruments commonly used for assessing violence risk, such as the PCL: SV.

*Tolman & Mullendore (2003)*

Tolman and Mullendore conducted a postal mail questionnaire with a group of clinical psychologists and a group of forensic psychologists in order to compare their violence risk assessment practices. The first group consisted of 200 randomly selected psychologists licensed to practice in the state of Michigan, and the second was composed of 182 diplomates of the American Board of Forensic Psychology. The researchers found the most commonly used instruments in the risk assessment process by clinical psychologists were the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 1994), the MMPI-2, and the Rorschach Inkblot Method (Exner et al., 2008), whereas the top three most commonly used instruments by forensic psychologists were the MMPI-2, the DSM-IV and the PCL-R. The researchers argued that board-certified forensic psychologists are more capable of providing a relevant and empirically-based foundation to assist triers of fact in making important riskrelated decisions, as they are better able to educate legal professionals on the intended use and findings of specialized risk assessment tools.

*Viljoen et al. (2010)*



Viljoen and colleagues conducted a Web-based survey of 130 psychologists with various training backgrounds (mostly clinical or counselling psychology) to examine the use of violence risk assessment tools in forensic evaluations of juveniles and adults. The survey was disseminated to the members of five organizations: American Board of Forensic Psychology, American Psychology-Law Society, the International Association of Forensic Mental Health Services, the Criminal Justice Section of the Canadian Psychological Association, and the American College of Forensic Psychology. The researchers found that clinical psychologists were more likely to use a structured instrument in adult violence risk assessment than with juveniles. In adult violence risk assessments, the most commonly used tools were the Psychopathy Checklist measures (PCL-R and PCL:SV), the HCR-20, and the MMPI-2. In juvenile violence risk assessments, the most commonly used tools were the Wechsler Intelligence Scales (WAIS-III; Wechsler, 1997; WASI; Wechsler, 1999; WISC-IV; Wechsler, 2003), the MMPI-2, and the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2003). When opinions on SPJ versus actuarial instruments were requested, the majority of clinicians reported both could be useful. However, this was a global rating of usefulness and not related to any specific task (i.e., risk assessment, risk management, or risk monitoring). The survey also included questions regarding perceived challenges when conducting risk assessment. The most commonly reported challenge was the difficulty in obtaining records and collateral information to properly conduct a risk assessment. Further, when comparing the practices of older and younger clinicians, younger clinicians were found to be more likely to use structured risk assessment tools when evaluating adults.

**Table 5: Characteristics Concerning the Administration of Nine Clinical Surveys of Structured Violence Risk Assessment**

Administration Characteristic	Archer et al. (2006)	Bengtson & Pedersen (2008)	Green et al. (2010)	Higgins et al. (2005)	Hawley et al. (2010)	Khiroya et al. (2009)	Lally (2003)	Tolman & Mullendore (2003)	Viljoen et al. (2010)
Language of survey	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear
No. of professional disciplines sampled	1	2	5	1	3	1	1	1	1
Method of administration	Web-based	Mail	Web-based	Mail	Handed out	Mail	Mail	Mail	Web-based
No. of sources included	2 (AP-LS list & ABFP diplomate directory)	Unstated/ Unclear	1 (CFMHS directory)	1 (Bindman et al [1999] database)	1 (Hertfordshire NHS Trust) <sup>1</sup>	1 (Internet database)	2 (ABFP directory, ABFP Web-site)	2 (ABFP directory & MBP database)	5 (AP-LS, IAFMHS, ABFP, ACFP & CPA-CJS directories)
No. of reminders sent	Unstated/ Unclear	1	Unstated/ Unclear	2	Unstated/ Unclear	2	1	Unstated/ Unclear	2
Incentives offered to respondents	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear
Queried use in RA, RMan, and/or RMon	RA	RA + RMan	RA	RA	RA	RA	RA	RA	RA
Perceived usefulness measured	No	No	No	No	Yes	Yes	No	No	Yes

*Note.* RA = risk assessment; RMan = risk management; RMon = risk monitoring; AP-LS = American Psychology-Law Society; ABFP = American Board of Forensic Psychology; CFMHS = Community Forensic Mental Health Services; MBP = Michigan Board of Psychology; IAFMHS = International Association of Forensic Mental Health Services; ACFP = American College of Forensic Psychology; CPA-CJS = Canadian Psychological Association – Criminal Justice Section.

<sup>1</sup>Excluding mental handicap, tertiary services, and secure or forensic services (Hawley et al., 2010).

**Table 6:** Characteristics Concerning the Findings of Nine Clinical Surveys on Structured Violence Risk Assessment

Results Characteristic	Archer et al. (2006)	Bengtson & Pedersen (2008)	Green et al. (2010)	Higgins et al. (2005)	Hawley et al. (2010)	Khiroya et al. (2009)	Lally (2003)	Tolman & Mullendore (2003)	Viljoen et al. (2010)
No. of respondents	152	41	10	45	300	29	64	164	130
Response rate (%)	56%	48%	83%	68%	50%	62%	35%	50%	50%
No. of countries surveyed	1	1	1	1	1	1	Unstated/ Unclear	1	Unstated/ Unclear
Mean age of respondents ( <i>SD</i> )	49.2 (12.0)	47.7 (10.4)	Unstated/ Unclear	Unstated/ Unclear	43 (Range = 23–62)	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	46.3 (12.1)
Men ( <i>N</i> ,%)	92 (60.5)	17 (42.5)	Unstated/ Unclear	Unstated/ Unclear	99 (33.0)	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	119 (59.8)
Mean no. of RA over lifetime ( <i>SD</i> )	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	429.5 (970.8)
Mean no. of RA using structured tools over lifetime ( <i>SD</i> )	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear
Mean no. of RA over last 12 months ( <i>SD</i> )	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	29.3 (39.4)
Mean no. of RA using structured tools over last 12 months	Unstated/ Unclear	Unstated/ Unclear	Median = 55	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear
Three most commonly used tools	1. PCL-R <sup>1</sup> 2. HCR-20 3. VRAG	1. PCL-R 2. PCL:SV 3. VRAG/ HCR-20	Unstated/ Unclear	Unstated/ Unclear	Unstated/ Unclear	1. PCL-R 2. HCR-20	Unstated/ Unclear	1. MMPI-2 2. DSM-IV 3. PCL-R <sup>2</sup>	1. PCL-R <sup>1</sup> 2. HCR-20 3. MMPI-2

*Note.* *N* = number of respondents; *M* = mean; *SD* = standard deviation; RA = risk assessment; HCR-20 = Historical Clinical Risk – 20 (Webster, Douglas, Eaves, & Hart, 1997); VRAG = Violence Risk Appraisal Guide (Harris, Rice, & Quinsey, 1993); MMPI-2 = Minnesota Multiphasic Personality Inventory – 2 (Butcher et al., 1989); DSM-IV = Diagnostic and Statistical Manual for Mental Disorders (4th edition; APA, 1994). PCL-R = The Psychopathy Checklist-Revised (Hare, 2003); PCL:SV = The Psychopathy Checklist - Screening Version (Hart, Cox, & Hare, 1995).

<sup>1</sup> = Inclusive of PCL:SV.

<sup>2</sup> = Most frequently used by forensic psychologists, diplomates of ABFP.

### 2.1.5 Discussion

The goal of the present study was to conduct a systematic review of surveys of violence risk assessment tool use and their perceived utility in practice to investigate the transparency of survey methodology and the consistency of findings. We identified nine surveys published in peer-reviewed journals between January 1, 2000 and January 1, 2013. The surveys ranged in focus: from the use of specific risk assessment tools at the unit-level to use of any forensic assessment tools at the level of the individual practitioner. A standardized checklist for survey research was used to assess the reporting quality of each survey, and supplementary coding sheets developed to extract characteristics regarding survey administration and results relevant to violence risk assessment.

There were three main findings of the present review. First, only half of the information needed for a transparent description of survey methodology was reported. This makes it difficult to compare the quality of findings across surveys. For example, only one study offered a clear description of the pretesting process of their survey. This makes it difficult to compare and assess the quality of the research measures used. Second, analysis of survey characteristics revealed considerable variation in response rates, suggesting a need for a more standardized method of dissemination and questionnaire administration. Third, and arguably most importantly, it appears that there is considerable variation in the extent of risk assessment tool use across countries and professional disciplines, ranging from 19% (Bengston & Pedersen, 2008) to 82% (Lally, 2003). This is a promising start, but given consistent findings that structured assessments outperform unstructured clinical judgments (Hilton, Harris, & Rice, 2006), there is a continued need for knowledge dissemination and training in the use of assessment tools.

### **Implications**

The findings of the systematic review may have important implications for both researchers and practitioners. Given our finding that there is a need for greater transparency in the violence risk assessment survey literature, we recommend that future research reports follow a standardized reporting guideline such as the checklist developed by Bennet and colleagues (2011) or the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) Statement (Eysenbach, 2004) and should endeavor to fulfill as many of the checklist criteria as possible to maximize transparency. A second important research implication emerges from the fact that there was a high variability in response rate between surveys (35% to 83%). This variability could be due to paper-and-pencil surveys disseminated via postal mail having higher response rates than Web-based surveys (Shih & Fan, 2009). To maximize response rates in future surveys, researchers are advised to use accepted “gold standard” approaches to survey design and dissemination, such as the Dillman Total Design Method (Dillman, Smyth, & Christian, 2009). The method consists of following specific steps in disseminating the questionnaire, from sending potential participants a cover letter which describes the purpose of the study, to specific patterns of follow-up. Regarding clinical implications, the findings across surveys suggest that practitioners and administrators interested in implementing widely-used and accepted risk assessment instruments may wish to consider SPJ tools such as the HCR-20 and actuarial tools such as the VRAG. Which approach to risk assessment is preferred is largely dependent upon the intended use of the tool, with SPJ instruments being more useful in the risk formulation and risk management process at the expense of introducing potential human judgment biases, whereas actuarial instruments are simpler to administer but rely upon probabilistic estimates of risk determined at the group rather than the individual-level (Hart & Cooke, 2013; Singh, 2013). Additional practical considerations include the fixed and variable costs associated with instruments (e.g., costs of manuals and coding sheets) and agency-specific needs and constraints.

**Future Directions**

Our review found that relatively few surveys have been published on the use of risk assessment tools in practice, despite the growing demand for reliable implementation of these instruments in forensic mental health practice. Hence, future research using qualitative and mixed-method approaches such as surveying is warranted. Large, cross-jurisdictional surveys that target multiple professional disciplines are particularly needed. We found no studies surveying the average number of risk assessments conducted using structured tools by respondents over their lifetime, whether they offered incentives to participants or not, or the perceived utility of risk assessment tools for risk management or risk monitoring. Future research may wish to develop surveys targeting these gaps in the current knowledge about practitioners' opinions on risk assessment tool use. In addition, future surveys may wish to focus on differences in the perceived utility of instruments for risk assessment, management, and monitoring. Also, the role of clinical judgement in the administration and interpretation of actuarial tools should be examined.

**Limitations**

There were several limitations to the present review. First, we only included surveys of general violence risk

assessment tools, meaning we excluded alternative areas of forensic risk assessment like sex offender and general recidivism risk assessment (e.g., Jackson & Hess, 2007; Taxman et al., 2007). Second, we did not focus on other aspects of practice relevant to risk assessment such as investigating the perceived importance of individual risk factors for violence (e.g., Elbogen et al., 2002) or examining case law on the role and relevance of the PCL-R in court settings (e.g., DeMatteo & Edens, 2006). Third, only English-based surveys conducted since the year 2000 were included, hence excluding studies written in other languages or published before this period (e.g., Lees-Haley, Smith, Williams, & Dunn, 1996). Fourth and finally, we

only included published studies, omitting unpublished surveys (e.g., Foellmi, Rosenfeld, Rotter, Greenspan, Khadivi, 2013).

### **Conclusion**

A number of structured tools have been developed over the past several decades to assess risk for future violence and there now is a large body of literature investigating the predictive validity of such tools and controversies surrounding them (e.g., Hart & Cooke, 2013; Heilbrun, Douglas, & Yasuhara, 2009; Skeem & Monahan, 2011; Troquete, van den Brink, Beintema, Mulder, van Os, & Schoevers, 2013). Our review found that, relative to the large psychometric literature on risk assessment tools, few studies have been published surveying the use of risk assessment tools in practice. This despite the growing demand for reliable implementation of these instruments in forensic mental health practice. Similar to quantitative research in forensic risk assessment (cf. Singh, Desmarais, & Van Dorn, 2013), there is a continued need for high quality surveys into the use and perceived utility of violence risk assessment in practice.

- 2.2 **Singh, J. P.**, Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., de Ruitter, C., Dean, K., Doyle, M., Folino, J., Godoy, V., Grann, M., Ho, R., Large, M., Nielsen, L. H., Pham, T., Rebocho, R. M., Reeves, K., Rettenberger, M., & Otto, R. K. (2014). Use and perceived utility of structured violence risk assessment tools in 44 countries: Findings from the IRiS Project. *International Journal of Forensic Mental Health, 13*, 193-206.

### 2.2.1 Abstract

Mental health professionals are routinely called upon to assess the risk of violence presented by their patients. Prior surveys of risk assessment methods have been largely circumscribed to individual countries and have not compared the practices of different professional disciplines. Therefore, a Web-based survey was developed to examine methods of violence risk assessment across six continents, and to compare the perceived utility of these methods by psychologists, psychiatrists, and nurses. The survey was translated into nine languages and distributed to members of 59 national and international organizations. Surveys were completed by 2135 respondents from 44 countries. Respondents in all six continents reported using instruments to assess, manage, and monitor violence risk, with over half of risk assessments in the past 12 months conducted using such an instrument. Respondents in Asia and South America reported conducting fewer structured assessments, and psychologists reported using instruments more than psychiatrists or nurses. Feedback regarding outcomes was not common: respondents who conducted structured risk assessments reported receiving feedback on accuracy in under 40% of cases, and those who used instruments to develop management plans reported feedback on whether plans were implemented in under 50% of cases. When information on the latter was obtained, risk management plans were not implemented in over a third of cases. Results suggest that violence risk assessment is a global phenomenon, as is the use of instruments to assist in this task. Improved feedback following



risk assessments and the development of risk management plans could improve the efficacy of health services.

### 2.2.2 Introduction

In light of heightened media attention on the link between violence and mental illness, there has been an increased demand for accurate and reliable methods of assessing violence risk (Brown, 2013). This focus on prevention is not new, however. The World Health Organization named violence prevention as one of its priorities over a decade ago (WHO, 2002). Moreover, current clinical guidelines for psychologists (American Psychological Association, 2006), psychiatrists (American Psychiatric Association, 2004; National Institute for Health and Clinical Excellence, 2009), and nurses (Nursing and Midwifery Council, 2004) recommend the routine assessment of violence risk for patients diagnosed with major mental illnesses. In recent decades, numerous violence risk assessment instruments have been developed to aid in this task. These instruments combine known risk and protective factors for violence either mechanically (the “actuarial approach”) or based on clinical discretion (the “structured professional judgment”, or SPJ, approach). They have been widely implemented in mental health and criminal justice settings, where they are used by psychologists, psychiatrists, and nurses to inform medico-legal decisions including commitment, classification, service plan development, and release (Conroy & Murrie, 2007).

As there are a large number of risk assessment tools available, practitioners are faced with the challenge of selecting the instrument that they feel to be the best fit for their population and that will best guide treatment planning. Indeed, recent meta-analyses suggest that risk assessment instruments may discriminate between violent and non-violent individuals with comparable accuracy (Yang, Wong, & Coid, 2010), implying that it may not be possible to base tool choice solely on predictive validity. In light of such findings, experts have recommended a shift in focus during the tool selection process (Skeem & Monahan,

2011), concentrating on the assessment needs of the practitioner in terms of the purpose of the evaluation, the population being assessed, and the outcome of interest (Singh, Grann, & Fazel, 2011). Thus, knowledge of which tools are currently being used in practice and which of them colleagues working in similar settings believe to be most useful may be informative. Surveys represent one approach to obtaining such information.

According to a search of PsycINFO, EMBASE, and MEDLINE, nine surveys have been published between January 1, 2000 to January 1, 2013 investigating violence risk assessment practices (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Bengtson & Pedersen, 2008; Green, Carroll, & Brett, 2010; Hawley, Gale, Sivakumaran, & Littlechild, 2010; Higgins, Watts, Bindman, Slade, & Thornicroft, 2005; Khiroya, Weaver, & Maden, 2009; Lally, 2003; Tolman & Mullendore, 2003; Viljoen, McLachlan, & Vincent, 2010). The studies have provided evidence that risk assessment tools are commonly used in practice by psychologists in the United States, the United Kingdom, Denmark, and Australia. Though the quality of these surveys vary (Hurducas, Singh, de Ruiters, & Petrila, in this issue), they have consistently found that actuarial instruments and personality scales are used more commonly in the violence risk assessment process than SPJ instruments.

These surveys have advanced our understanding of the use of violence risk assessment tools, but also share important limitations. First, no surveys have been published comparing what instruments are used in routine practice on different continents. Second, previous surveys have not compared patterns of tool use and perceived utility across professional disciplines. Third, previous surveys have not attempted to disentangle risk assessment, management, and monitoring practices. Consequently, many questions remain regarding the application of risk assessment tools in practice. Specifically, what instruments are currently being used, how frequently, in what context, by whom, and where? The answers to such questions may help guide individual clinicians working with mental health and criminal

justice populations to identify and implement the risk assessment tools with the greatest acceptability, efficacy, and fidelity (Andrews & Bonta, 2010). Hence, the present study aimed to investigate violence risk assessment practices in psychologists, psychiatrists, and nurses on six continents using a multilingual Web-based survey.

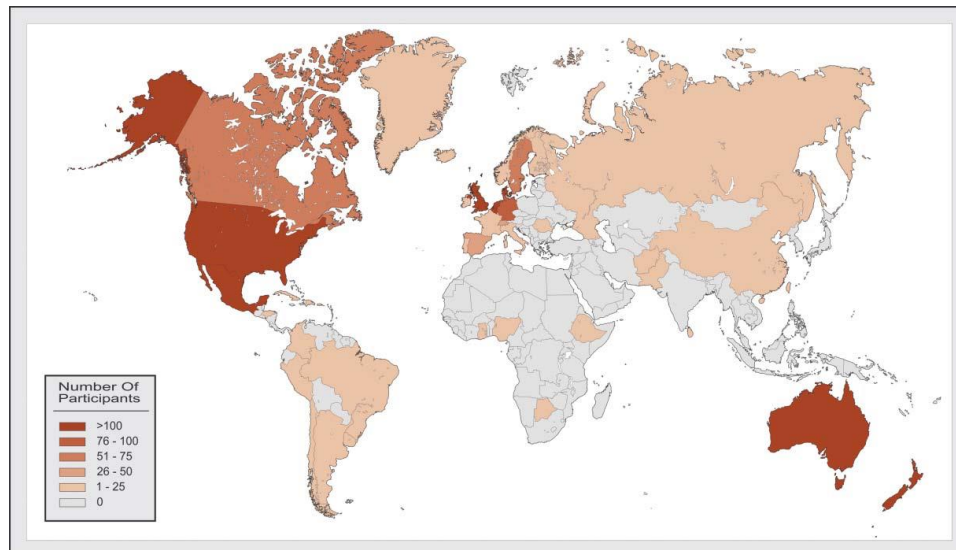
### 2.2.3 Methods

#### **Respondents**

Mental health professionals were eligible to participate if they were between the ages of 18 to 65 years and had assessed the violence risk of at least one adult in their life-time (N = 2135). Respondents included psychologists (n = 889, 41.6%), psychiatrists (n = 368, 17.2%), nurses (n = 622, 29.1%), and other professionals (n = 256, 12.0%) in 44 countries (Figure 2). The majority of respondents were from Europe (n = 1062, 49.7%) followed by North America (n = 444, 20.8%), Australasia (n = 112, 5.3%), Asia (n = 60, 2.8%), South America (n = 57, 2.7%), and Africa (n = 4, 0.2%). Demographic and clinical characteristics by professional discipline and continent are provided in Table 7 and Table 8, respectively, and for the overall sample in Appendix 1.

#### **Survey**

The survey included closed-ended questions developed through a review of the violence risk literature and drawn from previous surveys of clinicians concerning forensic assessment practices. Questions were organized into three blocks: (1) demographic and clinical characteristics, (2) prevalence and frequency of risk assessment instrument use, and (3) use and perceived utility of instruments in risk assessment, management, and monitoring. In the first block, respondents were asked about their demographic backgrounds and clinical activities over the past 12 months. Specifically, respondents were asked to approximate the total number of violence risk assessments conducted over their lifetime as well as in the past 12 months, estimating the percentage of those assessments conducted with the aid of an



**Figure 2:** Countries participating in an international survey on violence risk assessment practices.

instrument. Respondents also reported how often they received feedback concerning the accuracy of their risk assessments, as well as how often they learned whether the risk management plans they developed were implemented. In the second block, respondents reported the prevalence and frequency with which they used specific instruments in the risk assessment process over the past 12 months. (A list of instruments was constructed using recent reviews of the risk assessment literature, and respondents could identify up to three additional measures.) Frequency of use was rated on a 6-point Likert-type scale (0 Almost never; 5 Always). In the third block, respondents reported the tasks for which they used the specific tools identified in the second block (i.e., to inform judgments of violence risk, to develop violence risk management plans, and/or to monitor such plans). Perceived utility of instruments in the identified task(s) was rated on a 7-point Likert-type scale (0 = Very useless; 6 = Very useful).

**Table 7:** Demographic and Clinical Characteristics of Survey Respondents by Continent

Characteristic	Continent of Practice Over the Past 12 Months <sup>a</sup>									
	North America (n = 444)		South America (n = 57)		Europe (n = 1062)		Asia (n = 60)		Australasia (n = 112)	
<b>Demographic</b>										
Men (n, %)	164	(36.94)	30	(52.63)	431	(40.58)	22	(36.67)	43	(38.39)
Age in years (M, SD)	46.13	(11.91)	43.03	(9.36)	43.12	(10.55)	37.56	(10.14)	45.29	(10.35)
Years in practice (M, SD)	17.59	(11.62)	16.05	(10.55)	14.92	(10.20)	11.94	(9.45)	19.26	(11.54)
<b>Clinical setting over past 12 months</b>										
General hospital (M <sub>% Time</sub> , SD)	18.09	(34.09)	10.83	(22.07)	2.67	(12.50)	10.61	(26.10)	10.92	(25.96)
Private practice (M <sub>% Time</sub> , SD)	27.80	(38.19)	24.52	(29.29)	7.51	(21.45)	6.33	(24.00)	19.61	(33.22)
Non-forensic psych hospital (M <sub>% Time</sub> , SD)	6.97	(21.81)	5.65	(15.55)	16.14	(32.69)	28.40	(39.30)	9.18	(23.17)
Non-forensic psych clinic (M <sub>% Time</sub> , SD)	7.71	(23.08)	1.87	(9.82)	12.93	(29.55)	15.53	(30.91)	15.86	(31.69)
Forensic psych hospital (M <sub>% Time</sub> , SD)	9.08	(25.82)	4.74	(14.97)	24.52	(39.31)	19.58	(34.97)	6.74	(21.72)
Forensic psych clinic (M <sub>% Time</sub> , SD)	5.09	(18.21)	6.12	(19.76)	9.91	(25.39)	2.76	(10.86)	6.52	(19.71)
Correctional institute (M <sub>% Time</sub> , SD)	8.29	(24.09)	20.25	(36.12)	15.31	(32.39)	13.85	(32.09)	18.03	(35.29)
Other (M <sub>% Time</sub> , SD)	1.50	(9.51)	11.83	(26.68)	2.12	(11.45)	0.16	(1.29)	4.73	(17.40)
<b>Professional responsibilities over past 12 months</b>										
Practice (M <sub>%</sub> , SD)	57.08	(30.88)	54.38	(23.62)	48.63	(25.30)	68.36	(19.09)	59.09	(24.80)
Administrative duties (M <sub>%</sub> , SD)	16.57	(18.82)	15.73	(17.02)	24.89	(17.00)	12.31	(11.68)	18.38	(16.82)
Teaching or supervision (M <sub>%</sub> , SD)	14.14	(16.56)	13.05	(11.74)	11.85	(12.02)	9.56	(8.43)	14.38	(17.43)
Research (M <sub>%</sub> , SD)	8.47	(15.35)	6.40	(8.90)	6.23	(13.21)	7.71	(10.49)	6.33	(10.14)
Other (M <sub>%</sub> , SD)	3.71	(13.65)	10.42	(21.89)	8.41	(19.39)	2.03	(9.03)	1.82	(6.64)
<b>Risk assessment history</b>										
RA over lifetime (M, SD)	573.47	(1495.54)	701.98	(1655.74)	413.28	(1914.41)	364.40	(665.50)	841.23	(2735.87)
RA with SRAI over lifetime (M <sub>%</sub> , SD)	52.13	(38.71)	40.22	(34.50)	58.88	(37.85)	33.20	(36.04)	62.08	(35.93)
RA over past 12 months (M, SD)	42.44	(95.01)	50.39	(77.02)	36.12	(82.29)	78.35	(175.55)	51.95	(120.45)
RA with SRAI in past 12 months (M <sub>%</sub> , SD)	51.24	(42.92)	41.66	(37.02)	63.04	(40.75)	30.20	(37.91)	62.80	(42.17)
<b>Characteristics of examinees over past 12 months</b>										
Men (M <sub>%</sub> , SD)	63.14	(37.07)	72.89	(31.61)	80.13	(27.74)	69.40	(31.27)	80.27	(27.60)
Psychotic disorder (M <sub>%</sub> , SD)	24.27	(30.48)	16.52	(23.47)	31.60	(32.33)	50.26	(30.79)	36.17	(34.68)
Mood disorder (M <sub>%</sub> , SD)	28.34	(28.22)	14.33	(17.11)	10.02	(16.67)	20.73	(26.91)	22.26	(25.20)
Anxiety disorder (M <sub>%</sub> , SD)	17.20	(24.35)	7.96	(14.50)	6.19	(14.61)	7.26	(18.45)	10.56	(17.43)
SU disorder (M <sub>%</sub> , SD)	35.37	(33.33)	43.82	(30.88)	27.39	(30.50)	25.15	(28.87)	43.88	(35.50)
Personality disorder (M <sub>%</sub> , SD)	28.91	(33.02)	35.01	(32.26)	41.19	(31.77)	24.08	(27.77)	37.02	(28.93)
Other disorder (M <sub>%</sub> , SD)	9.58	(22.13)	11.22	(21.52)	11.56	(23.75)	6.78	(15.43)	6.84	(19.68)

Note. n = number of respondents; M = mean; SD = standard deviation; Psych = psychiatric; SU = substance use; SRAI = structured risk assessment instrument.

<sup>a</sup>Excluding respondents from Africa (n = 4).

## Procedure

The study was conducted in four phases between January to December 2012: (1) material development, (2) translation, (3) distribution, and (4) data analysis. The institutional review board at the University of South Florida approved all study procedures and waived the need for written informed consent (IRB Approval Number: Pro00007104).

**Table 8:** Demographic and Clinical Characteristics of Survey Respondents by Professional Discipline

Characteristic	Professional Discipline <sup>a</sup>					
	Psychology ( <i>n</i> = 889)		Psychiatry ( <i>n</i> = 368)		Nursing ( <i>n</i> = 622)	
Demographic						
Men ( <i>n</i> ,%)	321	(36.11)	208	(56.52)	225	(36.17)
Age in years ( <i>M</i> , <i>SD</i> )	41.70	(11.32)	46.96	(10.30)	46.00	(9.94)
Years in practice ( <i>M</i> , <i>SD</i> )	13.20	(9.78)	16.83	(9.72)	20.28	(11.30)
Clinical setting over past 12 months						
General hospital ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	3.03	(14.65)	10.50	(23.42)	13.57	(31.16)
Private practice ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	19.60	(33.94)	16.35	(29.52)	6.32	(21.84)
Non-forensic psych hospital ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	5.47	(19.63)	22.65	(34.25)	21.10	(36.69)
Non-forensic psych clinic ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	6.17	(21.06)	13.94	(28.03)	20.93	(36.69)
Forensic psych hospital ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	17.94	(35.18)	20.03	(34.20)	18.22	(36.18)
Forensic psych clinic ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	9.10	(24.72)	7.41	(18.96)	4.70	(18.45)
Correctional institute ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	23.44	(38.62)	4.96	(14.78)	2.46	(13.30)
Other ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	3.22	(14.70)	0.87	(6.32)	0.67	(5.02)
Professional responsibilities over past 12 months						
Practice ( <i>M</i> <sub>%</sub> , <i>SD</i> )	51.22	(26.83)	61.76	(22.23)	45.66	(29.39)
Administrative duties ( <i>M</i> <sub>%</sub> , <i>SD</i> )	19.82	(16.14)	17.49	(17.13)	27.53	(19.90)
Teaching or supervision ( <i>M</i> <sub>%</sub> , <i>SD</i> )	13.90	(15.08)	11.06	(8.94)	13.07	(15.28)
Research ( <i>M</i> <sub>%</sub> , <i>SD</i> )	8.87	(15.90)	6.01	(9.75)	4.28	(10.54)
Other ( <i>M</i> <sub>%</sub> , <i>SD</i> )	6.20	(16.38)	3.66	(12.39)	9.45	(21.12)
Risk assessment history						
RA over lifetime ( <i>M</i> , <i>SD</i> )	207.90	(690.72)	624.37	(1791.65)	650.05	(2401.35)
RA with SRAI over lifetime ( <i>M</i> <sub>%</sub> , <i>SD</i> )	67.35	(36.61)	36.49	(35.31)	48.88	(37.56)
RA over past 12 months ( <i>M</i> , <i>SD</i> )	21.99	(52.77)	45.62	(95.22)	47.75	(119.58)
RA with SRAI in past 12 months ( <i>M</i> <sub>%</sub> , <i>SD</i> )	72.62	(37.56)	43.84	(40.65)	48.35	(42.26)
Characteristics of examinees over past 12 months						
Men ( <i>M</i> <sub>%</sub> , <i>SD</i> )	84.33	(27.99)	76.76	(27.89)	65.64	(31.75)
Psychotic disorder ( <i>M</i> <sub>%</sub> , <i>SD</i> )	20.56	(29.46)	45.83	(30.92)	38.07	(32.13)
Mood disorder ( <i>M</i> <sub>%</sub> , <i>SD</i> )	14.92	(22.37)	13.84	(20.14)	17.15	(22.14)
Anxiety disorder ( <i>M</i> <sub>%</sub> , <i>SD</i> )	8.67	(17.51)	5.73	(14.71)	9.87	(18.21)
SU disorder ( <i>M</i> <sub>%</sub> , <i>SD</i> )	32.48	(32.80)	33.81	(31.70)	23.67	(28.64)
Personality disorder ( <i>M</i> <sub>%</sub> , <i>SD</i> )	45.03	(33.21)	38.51	(30.60)	29.09	(28.89)
Other disorder ( <i>M</i> <sub>%</sub> , <i>SD</i> )	12.31	(25.15)	8.72	(19.52)	8.28	(20.09)

Note. *n* = number of respondents; *M* = mean; *SD* = standard deviation; Psych = psychiatric; SU = substance use; SRAI = structured risk assessment instrument.

<sup>a</sup>Excluding respondents who self-identified as being members of other professional disciplines (*n* = 256).

In Phase 1 (January 2012–February 2012), the Web-based survey was constructed using Qualtrics electronic survey software ([www.Qualtrics.com](http://www.Qualtrics.com)). The list of survey questions were compiled in English and piloted by members of the Florida Mental Health Institute as well as 16 international experts representing the countries of Argentina, Australia and New Zealand, Belgium, Canada, Chile, Denmark, Germany, Hong Kong, Mexico, The Netherlands, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. These collaborators provided feedback that was used to make further refinements prior to translation and distribution.

In Phase 2 (March 2012–August 2012), the survey and participation letter were professionally translated from English into eight additional languages: Danish, Dutch, French, German, Portuguese, Spanish (Latin American), Spanish (European), and Swedish.

Translation services were provided by Software and Documentation Localization International ([www.SDL.com](http://www.SDL.com)). Translated materials were then sent to the international collaborators for back-translation. Identified discrepancies were corrected by the first author.

In Phase 3 (September 2012–November 2012), participation letters were distributed electronically via ListServes, membership directories, or bulletins of 59 national and international professional organizations (see Appendix 2 for a full list). The letters were distributed by each expert collaborator in their resident country's native language. Where available, the membership of at least three national organizations was targeted: (1) a national organization of psychologists (e.g., American Psychology-Law Society), (2) a national organization of psychiatrists (e.g., American Academy of Psychiatry and the Law), and (3) a national organization of nurses (e.g., Forensic Psychiatric Nurses Council). Where available, organizations of forensic specialists were identified. The membership of international forensic mental health organizations (e.g., International Association of Forensic Mental Health) was also targeted.

To the extent possible, survey distribution followed the Dillman Total Design Survey Method (Dillman, Smyth, & Christian, 2009). Specifically, participation letters were sent via e-mail on a Friday and contained direct and active links to the survey. Two reminder e-mails were sent in seven day increments after the initial distribution to remind potential respondents about the study. A fourth e-mail was also sent indicating a final opportunity to participate. Respondents who completed the survey and volunteered their e-mail addresses were entered into a raffle for eight cash prizes, each valued at \$50 USD. At the end of the data collection period, winners were randomly selected from the pool of respondents.

In Phase 4 (December 2012 to August 2013), respondent data was exported from Qualtrics to STATA/IC 10.1 and SPSS 17.01 for analysis. Descriptive and statistical analyses were conducted on the 12 most commonly used instruments in the violence risk assessment

process. However, over 200 commercially available instruments and a further 200 institutionally- or individually-developed instruments were reported as being used. Frequency distributions were examined and measures of central tendency and dispersion were calculated for all variables. Differences between continents (North American, South America, Europe, Asia, Australia)<sup>1</sup> and professional disciplines (psychologists, psychiatrists, nurses)<sup>2</sup> regarding the percentage of assessments conducted using an instrument and the regularity with which risk assessment and management feedback is given were explored via omnibus one-way ANOVAs. Statistical tests were two-tailed, and a Bonferroni-adjusted significance threshold of a  $\alpha = 0.004$  was used to address family-wise error due to multiple testing.

## 2.2.4 Results

### **Demographic and Clinical Characteristics**

The sample was composed of 2135 mental health professionals, the majority women ( $n = 1288$ , 60.3%). The average age of respondents was 43.9 years ( $SD = 11.0$ ), with an average of 15.9 years ( $SD = 10.7$ ) spent in practice. Approximately half of their time in the past 12 months was spent on clinical activities ( $M = 50.9\%$ ,  $SD = 28.2\%$ ), most often in forensic psychiatric hospitals ( $M = 17.5\%$ ,

$SD = 34.6\%$ ) followed by private practice ( $M = 15.0\%$ ,  $SD = 30.5\%$ ) and correctional institutions ( $M = 12.7\%$ ,  $SD = 29.9\%$ ). Additional professional responsibilities over the past 12 months included administrative duties ( $M = 22.0\%$ ,  $SD = 18.7\%$ ) and teaching ( $M = 13.2\%$ ,  $SD = 14.9\%$ ), with comparatively less time spent on research activities ( $M = 7.2\%$ ,  $SD = 14.4\%$ ).

---

<sup>1</sup> Given the small sample size from Africa, it was excluded from continental analyses.

<sup>2</sup> Professionals who did not self-report as being psychologists, psychiatrists, or nurses (e.g., social workers, counsellors, probation officer, law enforcement officer) were excluded from these analyses.



**Table 9:** Risk Assessment Instrument Prevalence and Frequency of Use Over the Past 12 Months by Continent

Instrument	Continent of Practice Over Past 12 Months <sup>a</sup>									
	North America (n = 286)		South America (n = 35)		Europe (n = 782)		Asia (n = 39)		Australasia (n = 112)	
	Number of Users (n, %)	Frequency of Use (M, SD)	Number of Users (n, %)	Frequency of Use (M, SD)	Number of Users (n, %)	Frequency of Use (M, SD)	Number of Users (n, %)	Frequency of Use (M, SD)	Number of Users (n, %)	Frequency of Use (M, SD)
COVR	44 (15.38)	3.27 (1.22)	4 (11.43)	2.66 (1.52)	11 (1.41)	3.11 (1.26)	3 (7.69)	4.00 (1.00)	1 (0.89)	3.00 (—)
FOTRES	9 (3.15)	4.11 (1.76)	0 (0.00)	— (—)	52 (6.65)	4.30 (1.76)	1 (2.56)	3.00 (—)	0 (0.00)	— (—)
HCR-20	102 (35.66)	4.58 (1.56)	14 (40.00)	4.21 (1.57)	499 (63.81)	4.43 (1.54)	18 (46.15)	3.16 (1.61)	44 (39.29)	4.14 (1.74)
HKT-30	1 (0.35)	6.00 (—)	0 (0.00)	— (—)	51 (6.52)	4.29 (1.71)	0 (0.00)	— (—)	0 (0.00)	— (—)
LSI-R	22 (7.69)	4.31 (1.78)	0 (0.00)	— (—)	37 (4.73)	3.64 (1.93)	1 (2.56)	4.00 (—)	18 (16.07)	4.17 (1.92)
PCL-R	101 (35.31)	4.21 (1.66)	18 (51.43)	3.83 (1.46)	366 (46.80)	3.77 (1.63)	10 (25.64)	2.90 (1.19)	30 (26.79)	2.52 (1.40)
PCL:SV	26 (9.09)	2.84 (1.43)	4 (11.43)	4.50 (1.29)	144 (18.41)	3.71 (1.67)	7 (17.95)	2.71 (1.70)	19 (16.96)	3.39 (1.72)
SAPROF	14 (4.90)	3.85 (1.79)	0 (0.00)	— (—)	125 (15.98)	3.68 (1.65)	1 (2.56)	2.00 (—)	5 (4.46)	3.20 (1.92)
START	29 (10.14)	3.50 (1.45)	0 (0.00)	— (—)	113 (14.45)	3.53 (1.78)	13 (33.33)	2.84 (1.95)	7 (6.25)	4.00 (1.26)
V-RISK-10	26 (9.09)	2.95 (1.16)	2 (5.71)	2.50 (0.70)	26 (3.32)	3.62 (1.68)	2 (5.13)	3.00 (1.41)	2 (1.79)	4.00 (—)
VRAG	47 (16.43)	4.23 (1.59)	2 (5.71)	3.00 (1.41)	123 (15.73)	3.79 (1.53)	2 (5.13)	4.00 (1.41)	6 (5.36)	1.75 (0.96)
VRS	50 (17.48)	3.16 (1.47)	5 (14.29)	3.20 (1.78)	29 (3.71)	2.76 (1.55)	4 (10.26)	4.50 (1.00)	20 (17.86)	4.80 (1.58)

Note. n = number of respondents; M = mean; SD = standard deviation; Users = number of respondents using instrument over past 12 months; Frequency = mean frequency of use rating over past 12 months; — = not applicable; COVR = Classification of Violence Risk (Monahan et al., 2005); FOTRES = Forensisch Operationalisiertes Therapie- und Risiko-Evaluations-System (Urbaniok, 2007); HCR-20 = Historical, Clinical, Risk Management-20 (Webster et al., 1997); HKT-30 = Historische, Klinische, Toekomstige-30 (Werkgroep Pilotstudy Risicotaxatie, 2002); LSI-R = Level of Service Inventory-Revised (Andrews & Bonta, 1995); PCL-R = Psychopathy Checklist-Revised (Hare, 2003); PCL:SV = Psychopathy Checklist: Screening Version (Hart et al., 1995); SAPROF = Structured Assessment of Protective Factors (de Vogel et al., 2007); START = Short-Term Assessment of Risk and Treatability (Webster et al., 2009); V-RISK-10 = Violence Risk Screening-10 (Hartvig et al., 2007); VRAG = Violence Risk Appraisal Guide (Quinsey et al., 2006); VRS = Violence Risk Scale (Wong & Gordon, 2009). Frequency use was measured using a six-point Likert scale (0 = *Almost never*; 5 = *Always*).

<sup>a</sup>Excluding respondents from Africa (n = 4).

### Risk Assessment Practices

Respondents reported conducting an average of 435.5 (SD = 1706.0) violence risk assessments in their lifetime, over half of which (M = 54.3%, SD = 38.9%) were conducted using a structured instrument. They conducted an average of 34.5 (SD = 86.9) violence risk assessments over the past 12 months, again over half of which (M = 58.3%, SD = 41.9%) were conducted using an instrument. Taking into consideration time spent conducting interviews, obtaining and reviewing records, and writing reports, structured violence risk

**Table 10:** Risk Assessment Instrument Prevalence and Frequency of Use Over the Past 12 Months by Professional Discipline

Instrument	Professional Discipline <sup>a</sup>					
	Psychology ( <i>n</i> = 737)		Psychiatry ( <i>n</i> = 255)		Nursing ( <i>n</i> = 345)	
	Number of Users ( <i>n</i> , %)	Frequency of Use ( <i>M</i> , <i>SD</i> )	Number of Users ( <i>n</i> , %)	Frequency of Use ( <i>M</i> , <i>SD</i> )	Number of Users ( <i>n</i> , %)	Frequency of Use ( <i>M</i> , <i>SD</i> )
COVR	18 (2.44)	3.06 (1.28)	10 (3.92)	2.40 (1.34)	19 (5.51)	3.78 (1.13)
FOTRES	27 (3.66)	4.44 (1.88)	16 (6.27)	4.18 (1.51)	13 (3.77)	4.23 (1.69)
HCR-20	379 (51.42)	4.64 (1.47)	141 (55.29)	4.06 (1.58)	112 (32.46)	4.08 (1.74)
HKT-30	46 (6.24)	4.21 (1.77)	2 (0.78)	5.50 (0.70)	2 (0.58)	5.00 (—)
LSI-R	54 (7.33)	4.01 (1.88)	7 (2.75)	2.00 (1.52)	1 (0.29)	6.00 (—)
PCL-R	363 (49.25)	3.90 (1.66)	99 (38.82)	3.58 (1.59)	23 (6.67)	2.60 (1.37)
PCL:SV	124 (16.82)	3.73 (1.60)	45 (17.65)	2.91 (1.66)	14 (4.06)	4.07 (1.77)
SAPROF	115 (15.60)	3.71 (1.63)	13 (5.10)	3.23 (1.64)	6 (1.74)	4.00 (2.28)
START	66 (8.96)	3.65 (1.70)	35 (13.33)	2.74 (1.44)	47 (13.62)	3.80 (1.87)
V-RISK-10	18 (2.44)	3.76 (1.52)	14 (5.49)	3.46 (1.45)	13 (3.77)	3.16 (1.40)
VRAG	122 (16.55)	3.85 (1.62)	34 (13.33)	3.97 (1.35)	9 (2.61)	3.37 (1.50)
VRS	44 (5.97)	3.83 (1.83)	13 (5.10)	2.84 (1.67)	36 (10.43)	3.27 (1.46)

Note. *n* = number of respondents; *M* = mean; *SD* = standard deviation; Users = number of respondents using instrument over past 12 months; Frequency = mean frequency of use rating over past 12 months; — = not applicable; COVR = Classification of Violence Risk (Monahan et al., 2005); FOTRES = Forensisch Operationalisiertes Therapie- und Risiko-Evaluations-System (Urbanik, 2007); HCR-20 = Historical, Clinical, Risk Management-20 (Webster et al., 1997); HKT-30 = Historische, Klinische, Toekomstige-30 (Werkgroep Pilotstudie Risicotaxatie, 2002); LSI-R = Level of Service Inventory-Revised (Andrews & Bonta, 1995); PCL-R = Psychopathy Checklist-Revised (Hare, 2003); PCL:SV = Psychopathy Checklist: Screening Version (Hart et al., 1995); SAPROF = Structured Assessment of Protective Factors (de Vogel et al., 2007); START = Short-Term Assessment of Risk and Treatability (Webster et al., 2009); V-RISK-10 = Violence Risk Screening-10 (Hartvig et al., 2007); VRAG = Violence Risk Appraisal Guide (Quinsey et al., 2006); VRS = Violence Risk Scale (Wong & Gordon, 2009). Frequency use was measured using a six-point Likert scale (0 = *Almost never*; 5 = *Always*).

<sup>a</sup>Excluding respondents who self-identified as being members of other professional disciplines (*n* = 256).

assessments over the past 12 months took an average of 7.8 hours (*SD* = 7.9) to conduct, whereas unstructured assessments took an average of 2.8 hours (*SD* = 2.7).<sup>3</sup>

Of those respondents who used instruments over the past 12 months, the majority used them for the purposes of risk assessment (*n* = 1134 of 1266 respondents who specified the purpose of their instrument use, 89.6%) followed by developing risk management plans (*n* = 869, 68.6%) and monitoring those plans (*n* = 499, 39.4%). Respondents who used instruments to structure their violence risk assessments reported receiving feedback on the accuracy of their assessments in an average of 36.5% (*SD* = 34.7%) of cases. Those who used instruments to develop risk management plans were made aware of whether those plans had been

<sup>3</sup> Findings concerning specific professional disciplines and continents are available upon request.

implemented in an average of 44.6% (SD = 34.7%) of cases. Where such information was available, respondents reported that their proposed management plans were implemented in an average of 65.4% (SD = 27.5%) of cases.

### **Comparisons by Geographic Location and Professional Discipline**

Analyses showed differences in the prevalence of instrument use as a function of geographic location and professional discipline. Compared to North America, Europe, and Australasia, respondents in Asia and South America reported completing a smaller proportion of risk assessments with the aid of an instrument both over the lifetime,  $F(4, 1706) = 11.06$ ,  $p < .001$ ,  $h^2 = 0.03$ , 95% CI [0.02, 0.05], as well as over the past 12 months,  $F(4, 1682) = 16.09$ ,  $p < .001$ ,  $h^2 = 0.04$ , 95% CI [0.02, 0.06]. In terms of professional discipline, psychologists reported using instruments to structure their violence risk assessments more often than did psychiatrists or nurses both over their lifetime,  $F(2, 1876) = 105.85$ ,  $p < .001$ ,  $h^2 = 0.10$ , 95% CI [0.07, 0.11] and in the past 12 months,  $F(2, 1503) = 82.35$ ,  $p < .001$ ,  $h^2 = 0.10$ , 95% CI [0.07, 0.11]. Nurses reported more often obtaining feedback on whether their risk management plans had been implemented,  $F(2, 770) = 10.04$ ,  $p < .001$ ,  $h^2 = 0.03$ , 95% CI [0.01, 0.05], and that their risk management plans were implemented more often than psychologists or psychiatrists,  $F(2, 660) = 10.19$ ,  $p < .001$ ,  $h^2 = 0.03$ , 95% CI [0.01, 0.06]. Finally, psychologists reported taking significantly longer to conduct both unstructured violence risk assessments,  $F(2, 202) = 10.06$ ,  $p < .001$ ,  $h^2 = 0.09$ , 95% CI [0.02, 0.12], and structured violence risk assessments,  $F(2, 896) = 57.33$ ,  $p < .001$ ,  $h^2 = 0.11$ , 95% CI [0.10, 0.16]. Specific Risk Assessment Instrument Use more than 200 different instruments were reported as being used in the violence risk assessment process, not including over 200 additional instruments developed for personal or institutional use only. In the present study, we describe the prevalence and perceived utility of those 12 instruments used most commonly

by respondents over the past year. Six of these were actuarial instruments and six were SPJ instruments.

The prevalence and frequency of risk assessment instrument use over the past 12 months is reported by professional discipline and continent in Tables 10 and 11, respectively, and for the overall sample in Appendix 3. Over both their lifetime and in the past 12 months, respondents reported that the instruments most commonly used in the violence risk assessment process were the Historical, Clinical, Risk Management-20 (HCR-20;  $n_{\text{Lifetime}} = 1032$  of 2135 respondents, 48.34%;  $n_{\text{Year}} = 669$  of 2135, 31.33%) (Webster, Douglas, Eaves, & Hart, 1997), Psychopathy Checklist-Revised (PCL-R;  $n_{\text{Lifetime}} = 836$ , 39.16%;  $n_{\text{Year}} = 513$ , 24.03%)<sup>4</sup> (Hare, 2003), and Psychopathy Checklist: Screening Version (PCL:SV;  $n_{\text{Lifetime}} = 409$ , 19.16%;  $n_{\text{Year}} = 195$ , 9.13%) (Hart, Cox, & Hare, 1995). Those who used specific instruments were also asked how frequently they used them. Respondents who used the HCR-20 ( $M = 3.71$ ,  $SD = 1.65$ ), PCL-R ( $M = 3.32$ ,  $SD = 1.58$ ), and the Historische, Klinische, Toekomstige-30 (HKT-30;  $M = 3.16$ ,  $SD = 1.73$ ) (Werkgroep Pilotstudie Risicotaxatie Forensische Psychiatrie, 2002) at some point in their lifetime reported using these most frequently. Over the past 12 months, the HCR-20 ( $M = 4.40$ ,  $SD = 1.58$ ), HKT-30 ( $M = 4.33$ ,  $SD = 1.71$ ), and the Forensisches Operationalisiertes Therapie-Risiko-Evaluations- System (FOTRES;  $M = 4.33$ ,  $SD = 1.71$ ) (Urbaniok, 2007) were the most frequently administered instruments by their users.

The HCR-20 was the instrument most commonly used for conducting violence risk assessments, developing risk management plans, and monitoring risk management plans (Table 11). Those who used SPJ instruments including the HCR-20, HKT-30, FOTRES, the Short-Term Assessment of Risk and Treatability (Webster, Martin, Brink, Nicholls, &

---

<sup>4</sup> Consistent with previous surveys on forensic risk assessment, we did not assume that the use of instruments that incorporate the PCL-R as an item necessarily meant that the PCL-R was used. For example, the HCR-20 authors have found that the scheme performs better without the PCL-R (Guy, Douglas, & Hendry, 2010) and the VRAG manual allows for prorating should this information be missing (Quinsey et al., 2006).

Desmarais, 2009), and the Structured Assessment of PROtective Factors (de Vogel, de Ruiter, Bouman, & de Vries Robb\_e, 2007) rated these tools, on average, as being very useful for these tasks. Notably, the HKT-30 and FOTRES were virtually only used by professionals practicing in Europe.

**Table 11:** Instrument Use in Violence Risk Assessment, Management, and Monitoring Over the Past 12 Months

Instrument	All respondents ( <i>N</i> = 976)					
	Number of RA Users ( <i>n</i> ,%)	Usefulness in RA ( <i>M</i> , <i>SD</i> )	Number of RMx Users ( <i>n</i> ,%)	Usefulness in RMx ( <i>M</i> , <i>SD</i> )	Number of RMon Users ( <i>n</i> ,%)	Usefulness in RMon ( <i>M</i> , <i>SD</i> )
COVR	37 (3.79)	3.81 (0.78)	33 (3.38)	3.73 (1.15)	28 (2.87)	3.96 (1.32)
FOTRES	50 (5.12)	4.14 (1.05)	35 (3.59)	4.49 (0.89)	28 (2.87)	4.36 (1.06)
HCR-20	588 (60.25)	4.44 (0.78)	453 (46.41)	4.40 (0.80)	237 (24.28)	4.13 (0.92)
HKT-30	46 (4.71)	4.52 (0.69)	31 (3.18)	4.48 (0.72)	20 (2.05)	4.00 (0.86)
LSI-R	66 (6.76)	4.09 (1.05)	51 (5.23)	3.90 (1.20)	22 (2.25)	3.59 (1.30)
PCL-R	461 (47.23)	4.26 (0.84)	326 (33.40)	3.75 (1.06)	162 (16.60)	3.09 (1.31)
PCL:SV	164 (16.80)	4.05 (0.89)	137 (14.04)	3.61 (0.99)	73 (7.48)	3.12 (1.29)
SAPROF	127 (13.01)	4.35 (0.83)	100 (10.25)	4.44 (0.73)	52 (5.33)	4.02 (0.98)
START	132 (13.52)	4.19 (0.97)	117 (11.99)	4.32 (0.88)	92 (9.43)	4.26 (0.85)
V-RISK-10	34 (3.48)	3.88 (0.77)	29 (2.97)	3.97 (1.09)	18 (1.84)	4.11 (1.08)
VRAG	151 (15.47)	4.03 (0.92)	97 (9.94)	3.21 (1.22)	55 (5.64)	2.93 (1.43)
VRS	72 (7.38)	4.14 (1.03)	69 (7.07)	4.10 (0.96)	42 (4.3)	4.10 (0.88)

*Note.* *n* = number of respondents; RA = risk assessment; RMx = risk management; RMon = risk monitoring; *M* = mean; *SD* = standard deviation. COVR = Classification of Violence Risk (Monahan et al., 2005); FOTRES = Forensisch Operationalisiertes Therapie- und Risiko-Evaluations-System (Urbanik, 2007); HCR-20 = Historical, Clinical, Risk Management-20 (Webster et al., 1997); HKT-30 = Historische, Klinische, Toekomstige-30 (Werkgroep Pilotstudie Risicotaxatie, 2002); LSI-R = Level of Service Inventory-Revised (Andrews & Bonta, 1995); PCL-R = Psychopathy Checklist-Revised (Hare, 2003); PCL:SV = Psychopathy Checklist: Screening Version (Hart et al., 1995); SAPROF = Structured Assessment of Protective Factors (de Vogel et al., 2007); START = Short-Term Assessment of Risk and Treatability (Webster et al., 2009); V-RISK-10 = Violence Risk Screening-10 (Hartvig et al., 2007); VRAG = Violence Risk Appraisal Guide (Quinsey et al., 2006); VRS = Violence Risk Scale (Wong & Gordon, 2009). Perceived utility was measured using a 7-point Likert scale (0 = *Very useless*; 6 = *Very useful*).

## Sensitivity Analysis

As a sensitivity analysis, univariate linear regression analyses were performed to investigate whether sex, age, or number of years in practice was associated with the percentage of risk assessments conducted using a structured instrument over respondents' lifetime and in the past 12 months. Respondent sex was not found to be associated with instrument use. Younger respondents were found to have conducted a higher percentage of

their assessments using structured instruments over their lifetime,  $t(2115) = 7.22$ ,  $p < .001$ ,  $b = 0.04$ , 95% CI [0.03, 0.06], as well as in the past 12 months,  $t(1676) = 3.94$ ,  $p < .001$ ,  $b = 0.03$ , 95% CI [0.01, 0.04]. Similarly, respondents earlier in their practice careers conducted a higher percentage of their assessments using structured instruments over their lifetime,  $t(2133) = 9.00$ ,  $p < .001$ ,  $b = 0.05$ , 95% CI [0.04, 0.06], as well as in the past 12 months,  $t(1687) = 5.74$ ,  $p < .001$ ,  $b = 0.04$ , 95% CI [0.02, 0.05].

### 2.2.5 Discussion

Despite the proliferation of violence risk assessment methods in mental health and criminal justice settings, research on what instruments are used in practice and their perceived utility is rare (Elbogen, Huss, Tomkins, & Scalora, 2005). Work comparing risk assessment procedures on different scarce, making it unclear whether clinicians working in different contexts should assume the generalizability of previous survey findings. Therefore, the present study aimed to survey the use and perceived utility of violence risk assessment methods in practice by 2135 psychologists, psychiatrists, and nurses on six continents. Respondents reported using over 400 instruments to assess, manage, and monitor violence risk, with over half of risk assessments in the past 12 months conducted using such an instrument. Due to the emphasis on assessment as part of their training and practice, the majority of respondents as well as the majority of tool users were psychologists.

The survey findings may have important implications for practice and research. First, the results identify which

structured instruments are being used by mental health professionals to conduct violence risk assessments, to inform the development of risk management plans and to assist in their monitoring. The findings also speak to the perceived utility of instruments in these tasks. This information may assist practitioners' selection of which risk assessment tools to implement. With the two leading approaches to structured risk assessment (actuarial and SPJ)

demonstrating similar popularity and with mechanical and clinically-based tools having similar reliability and accuracy (Fazel, Singh, Doll, & Grann, 2012), the focus of instrument selection should be on the goodness-of-fit between the population and setting in which a professional is working and those for which tools were designed. Additional practical considerations include administration time, cost, training needs, and personal preference for a tool's approach to assessment (Desmarais & Singh, 2013). Given that the users of SPJ instruments rated them as very useful in the development and monitoring of risk management plans, assessors working in rehabilitation and recovery-focused settings may wish to consider adopting such tools (e.g., HCR-20, HKT- 30, FOTRES, SAPROF, START). Instruments following this approach may be particularly useful internationally, as recent meta-analytic evidence suggests that probabilistic estimates of violence risk produced by actuarial risk assessment instruments may vary considerably depending on local factors (Singh, Fazel, Gueorguieva, & Buchanan, 2013, 2014). Findings also suggest that personality scales such as the Psychopathy Checklist measures continue to be used as part of the risk assessment process. Albeit such instruments may have an important role to play in developing responsive risk management plans, they have not been found to predict violence as accurately as tools explicitly designed for the purposes of violence risk assessment (Singh et al., 2011).

Second, findings suggest a need for increased communication about violence risk assessments. Respondents who used instruments to inform their assessments reported receiving any kind of feedback on their accuracy in only a third of cases (36.5%). However, social psychology research demonstrates that judgment accuracy increases when decision-makers receive feedback about their performance (Arkes, 1991). Therefore, violence risk assessors should be provided with follow-up information on their examinees whenever possible. This may be particularly helpful in the avoidance of false negative decisions, because individuals judged to be at higher risk will, in practice, be less likely to have access

to potential victims. We also found that respondents who used instruments to develop management plans frequently did not know whether their plans had been implemented (44.6%) and, amongst those who did, proposed plans were not implemented in over a third of cases (34.6%). The latter is of principal importance: what is the use of developing risk management plans if they are not implemented into practice? Risk assessments will not reduce violence unless their findings are communicated transparently and suggestions for risk management are executed (Heilbrun, Dvoskin, Hart, & McNiel, 1999). These findings require further research to clarify what feedback on risk assessments constituted and to what extent risk management plans were implemented. Which form of feedback (e.g., obtaining court records to view judges decisions, obtaining information from criminal registers, interviewing family members) is most effective in improving accuracy could be a promising area of future research.

Third, the results may inform the research agendas of several geographic regions. Fewer than half of risk assessments in South America and Asia over the past year were conducted with the use of a risk assessment tool, despite the large evidence base demonstrating the superiority of structured methods over unstructured clinical judgment. Though it may be that this continuing trend is due to cultural differences, it is also possible that more evidence of such superiority is needed using non-Western samples to be influential in practice. Additionally, despite an existent literature on the predictive validity of risk assessment tools in these regions (Folino, Marengo, Marchiano, & Ascazibar, 2004; Ho et al., 2013), the rarity of their use may also be due to a lack of familiarity with commercially available instruments or the unavailability of authorized translations (e.g., Telles, Day, Folino, & Taborda, 2009; Zhang, Chan, Cai, & Hu, 2012). Moving forward, clinical training programs in these areas may wish to incorporate modules on violence risk assessment tools, funding agencies may wish to issue grants to encourage the development of novel instruments



in native languages or the authorized translations of available tools, and there needs to be increased discussion in the field about the strengths and limitations of the contemporary literature and best-practice recommendations in international settings.

### **Limitations**

Limitations of the present study include coverage, sampling, and nonresponse errors characteristic of probability-based surveying methods (Couper, 2000), as well as both respondent- (e.g., lack of motivation, comprehension problems, reactivity) and software-related measurement error (e.g., technical difficulties). Specifically, a response rate was unable to be established for the present survey, impeding our ability to make a statement of the generalizability of our findings. Many of our respondents were members of more than one of the organizations that assisted in the dissemination process. Also, it is likely that some respondents heard about the survey through colleagues or friends, but may not necessarily have been members of the organizations sampled. These are limitations shared by previous Web-based surveys that have been disseminated using multiple ListServes (Archer et al., 2006; Viljoen et al., 2010). Future surveys should include as an item a list of the organizations through which they disseminated calls for participation. Respondents should be allowed to identify all those organizations of which they are members. Researchers can then request information from each organization as to its membership count for the date on which the calls for participation were made. This would allow statistical correction for overlap in organizational membership, and for the calculation of a response rate. It is also difficult to assess generalizability as information is not available regarding characteristics of nonrespondents, who may have differed systematically from respondents. For example, nonrespondents who employ violence risk assessment instruments may do so less often or have significantly poorer perceptions of their utility. Another issue of generalizability is evidence that men may be less likely than women to respond to surveys (Kwak & Radler,

2002; Underwood, Kim, & Matier, 2000). Hence, the current study findings should be treated as tentative until replication attempts are made in future research.

## **Conclusion**

The routine assessment of violence risk has become a global phenomenon, as has the use of instruments to assist in this task. Across continents, providing practitioners with feedback on the accuracy of their predictions and whether their management plans were implemented could improve the predictive validity of assessments as well as risk communication. Given the substantial evidence base supporting the benefits in reliability and validity of structured over unstructured assessment (Ægisdottir et al., 2006), the study of violence risk assessment methods in South America and Asia should be a public health research priority. And, as the prevalence of tool use grows in additional regions such as Africa (Roffey & Kaliski, 2012) and Eastern Europe (Jovanović et al., 2009), the importance of high-quality research into psychometric properties and fidelity in implementation will become ever more important.

### **2.2.6 Funding**

The research was funded by an American Psychology- Law Society Early Career Professional Grant. This sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

## 2.2.8 Appendix

**Demographic and Clinical Characteristics of All Survey Respondents**

Characteristic	All respondents ( <i>n</i> = 2135)	
<b>Demographic</b>		
Men ( <i>n</i> ,%)	847	(39.67)
Age in years ( <i>M</i> , <i>SD</i> )	43.93	(10.97)
Years in practice ( <i>M</i> , <i>SD</i> )	15.91	(10.71)
<b>Clinical setting over past 12 months</b>		
General hospital ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	8.66	(24.46)
Private practice ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	15.03	(30.46)
Non-forensic psych hospital ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	13.27	(29.62)
Non-forensic psych clinic ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	12.08	(28.59)
Forensic psych hospital ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	17.48	(34.59)
Forensic psych clinic ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	7.02	(21.35)
Correctional institute ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	12.68	(29.87)
Other ( <i>M</i> <sub>% Time</sub> , <i>SD</i> )	2.58	(13.07)
<b>Professional responsibilities over past 12 months</b>		
Practice ( <i>M</i> <sub>%</sub> , <i>SD</i> )	50.91	(28.23)
Administrative duties ( <i>M</i> <sub>%</sub> , <i>SD</i> )	21.95	(18.66)
Teaching or supervision ( <i>M</i> <sub>%</sub> , <i>SD</i> )	13.17	(14.86)
Research ( <i>M</i> <sub>%</sub> , <i>SD</i> )	7.18	(14.36)
Other ( <i>M</i> <sub>%</sub> , <i>SD</i> )	6.80	(17.79)
<b>Risk assessment history</b>		
RA over lifetime ( <i>M</i> , <i>SD</i> )	435.46	(1705.99)
RA with SRAI over lifetime ( <i>M</i> <sub>%</sub> , <i>SD</i> )	54.32	(38.93)
RA over past 12 months ( <i>M</i> , <i>SD</i> )	34.53	(86.87)
RA with SRAI in past 12 months ( <i>M</i> <sub>%</sub> , <i>SD</i> )	58.25	(41.94)
<b>Characteristics of examinees over past 12 months</b>		
Men ( <i>M</i> <sub>%</sub> , <i>SD</i> )	75.55	(31.87)
Psychotic disorder ( <i>M</i> <sub>%</sub> , <i>SD</i> )	32.29	(33.04)
Mood disorder ( <i>M</i> <sub>%</sub> , <i>SD</i> )	17.55	(23.54)
Anxiety disorder ( <i>M</i> <sub>%</sub> , <i>SD</i> )	10.22	(19.16)
SU disorder ( <i>M</i> <sub>%</sub> , <i>SD</i> )	31.14	(32.19)
Personality disorder ( <i>M</i> <sub>%</sub> , <i>SD</i> )	36.34	(32.26)
Other disorder ( <i>M</i> <sub>%</sub> , <i>SD</i> )	9.74	(22.27)

*Note.* *n* = number of respondents; *M* = mean; *SD* = standard deviation; Psych = psychiatric; SU = substance use; SRAI = structured risk assessment instrument.

**International and Intranational Organizations Involved in the Dissemination of Survey Materials**

1. Red Iberolatinoamericana de investigación y Docencia en Salud Mental Aplicada a lo Forense
2. Royal Australian and New Zealand College of Psychiatrists
3. SocieteRoyalededeMedecineMentalededeBelgique
4. Canadian Psychological Association - Criminal Justice Psychology Section
5. Gendarmeriade Chile
6. Dansk Psykologforening, Hospitals-Sektionen
7. Bundesfachvereinigung Leitender Krankenpflegepersonen der Psychiatrie e.V., Netzwerk Forensik
8. Hong Kong College of Psychiatrists
9. Colegio Nacional de Enfermeras
10. Nederlands Instituut van Psychologen - Forensic Psychology Section

11. Instituto Nacional de Medicina Legal
12. PSI-FORENSE ListServ
13. Swedish Medical Association
14. Schweizer Gesellschaft für Forensische Psychiatrie
15. Royal College of Nursing
16. American Academy of Psychiatry and the Law
17. International Association for Forensic Mental Health Services
18. Maestría en Salud Mental aplicada a lo Forense, Departamento de Postgrado, Facultad de Ciencias Médicas, Universidad Nacional de La Plata
19. Australian Psychological Society
20. Belgian College of Neuropsychopharmacology and Biological Psychiatry
21. Canadian Psychiatric Association
22. Dansk Retspsykologisk Selskab
23. Berufsverband Deutscher Psychologinnen und Psychologen e.V. (BDP), Sektion Rechtspsychologie
24. HongKong Psychological Society
25. Asociación Psiquiátrica Mexicana A.C.
26. Verpleegkundigen & Verzorgenden Nederland - Social Psychiatric Nurses Section
27. Guarda Nacional Republicana
28. Societat Catalana de Medicina Legal i Toxicologia Centre d'Estudis
29. Swedish Psychiatric Association
30. Schweizer Gesellschaft für Rechtspsychologie
31. Royal College of Psychiatrists
32. American Psychology-Law Society
33. American Institute for the Advancement of Forensic Studies
34. Australian College of Mental Health Nurses
35. Belgian Association for Psychological Sciences
36. Canadian Academy of Psychiatry and the Law
37. Psykologfagligt Forum, Øst
38. Deutsche Gesellschaft für Psychologie (DGPs), Fachgruppe Rechtspsychologie
39. Academy of Mental Health
40. Sociedad Mexicana de Psicología A.C.
41. Nederlandse Vereniging voor Psychiatrie
42. Direção Geral dos Serviços Prisionais
43. Jurídica Formación Especializada
44. Swedish Forensic Psychiatric Association
45. Schweizer Amt für Justiz
46. British Psychological Society
47. PSYLAWListServ
48. Association Francophone des Infirmières spécialisées en santé mentale et Psychiatrique
49. Dansk Psykiatrisk Selskab
50. Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN), Referat Forensische Psychiatrie
51. Policia Judiciária
52. Swedish Psychologists' Association

53. American Board of Forensic Psychology  
 54. Nationale Federatie van Belgische Verpleegkundigen Fédération National des  
 Infirmières de Belgique  
 55. Fagligt Selskab for Psykiatriske Sygeplejersker  
 56. Niedersächsisches Justizministerium, Abteilung Justizvollzug und Kriminologischer  
 Dienst  
 57. American Academy of Forensic Psychology  
 58. American Psychiatric Nurses Association  
 59. Forensic Behavioral Services, Inc.

### Risk Assessment Instrument Prevalence and Frequency of Use Over the Past 12 Months by All Survey Respondents

Instrument	All respondents ( <i>n</i> = 2135)	
	Number of Users ( <i>n</i> ,%)	Frequency of Use ( <i>M</i> , <i>SD</i> )
COVR	63 (2.95)	3.23 (1.21)
FOTRES	60 (2.81)	4.33 (1.71)
HCR-20	669 (31.33)	4.40 (1.58)
HKT-30	52 (2.44)	4.33 (1.71)
LSI-R	77 (3.61)	3.95 (1.88)
PCL-R	513 (24.03)	3.77 (1.66)
PCL:SV	195 (9.13)	3.56 (1.68)
SAPROF	144 (6.74)	3.66 (1.67)
START	160 (7.49)	3.50 (1.73)
V-RISK-10	54 (2.53)	3.29 (1.45)
VRAG	176 (8.24)	3.84 (1.57)
VRS	106 (4.96)	3.44 (1.68)

*Note.* *n* = number of respondents; *M* = mean; *SD* = standard deviation; Users = number of respondents using instrument over past 12 months; Frequency = mean frequency of use rating over past 12 months; — = not applicable; COVR = Classification of Violence Risk (Monahan et al., 2005); FOTRES = Forensisch Operationalisiertes Therapie- und Risiko-Evaluations-System (Urbanik, 2007); HCR-20 = Historical, Clinical, Risk Management-20 (Webster et al., 1997); HKT-30 = Historische, Klinische, Toekomstige-30 (Werkgroep Pilotstudy Risicotaxatie, 2002); LSI-R = Level of Service Inventory-Revised (Andrews & Bonta, 1995); PCL-R = Psychopathy Checklist-Revised (Hare, 2003); PCL:SV = Psychopathy Checklist: Screening Version (Hart et al., 1995); SAPROF = Structured Assessment of Protective Factors (de Vogel et al., 2007); START = Short-Term Assessment of Risk and Treatability (Webster et al., 2009); V-RISK-10 = Violence Risk Screening-10 (Hartvig et al., 2007); VRAG = Violence Risk Appraisal Guide (Quinsey et al., 2006); VRS = Violence Risk Scale (Wong & Gordon, 2009). Frequency use was measured using a six-point Likert scale (0 = *Almost never*; 5 = *Always*).

2.3 **Singh, J. P., Desmarais, S., & Van Dorn, R. A. (2013).** Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences & the Law*, 31, 55-73.

#### 2.3.1 Abstract

The objective of the present review was to examine how predictive validity is analyzed and reported in studies of instruments used to assess violence risk. We reviewed 47 predictive validity studies published between 1990 and 2011 of 25 instruments that were included in two recent systematic reviews. Although all studies reported receiver operating characteristic curve analyses and the area under the curve (AUC) performance indicator, this methodology was defined inconsistently and findings often were misinterpreted. In addition, there was between-study variation in benchmarks used to determine whether AUCs were small, moderate, or large in magnitude. Though virtually all of the included instruments were designed to produce categorical estimates of risk – through the use of either actuarial risk bins or structured professional judgments – only a minority of studies calculated performance indicators for these categorical estimates. In addition to AUCs, other performance indicators, such as correlation coefficients, were reported in 60% of studies, but were infrequently defined or interpreted. An investigation of sources of heterogeneity did not reveal significant variation in reporting practices as a function of risk assessment approach (actuarial vs. structured professional judgment), study authorship, geographic location, type of journal (general vs. specialized audience), sample size, or year of publication. Findings suggest a need for standardization of predictive validity reporting to improve comparison across studies and instruments.

#### 2.3.2 Introduction

Numerous risk assessment instruments have been introduced in recent decades. Recommended by current clinical guidelines for mental health professionals (Buchanan,

Binder, Norko, & Swartz, 2012; Department of Health, 2007; Nursing and Midwifery Council, 2004), these instruments are designed to aid in the assessment of risk for antisocial behavior, most commonly general violence, sexual violence, and criminal offending. In addition to their use in psychiatric settings, risk assessment tools are increasingly required by courts and correctional agencies (Skeem & Monahan, 2011). As assessments completed using these instruments inform medical and legal decisions of direct relevance to treatment, individual liberty and public safety (Janus, 2004; Monahan, 2006; Tyrer et al., 2010), research investigating their predictive validity is of considerable importance. Consequently, there is an abundance of studies evaluating the predictive validity of more than 150 risk assessment instruments (Singh, Serper, Reinharth, & Fazel, 2011). However, little is known about how predictive validity is described and interpreted in the risk assessment literature, including the factors that influence analytic and reporting practices.

### **Measuring Predictive Validity**

Performance indicators used in the risk assessment literature to measure predictive validity can generally be divided into three categories: (1) those that indicate the ability to accurately identify groups of individuals most likely to commit an antisocial act; (2) those that indicate the ability to accurately identify groups of individuals least likely to commit an antisocial act; and (3) those that indicate predictive abilities overall (Singh, Grann, & Fazel, 2011). Such distinctions are not inconsequential: the use of different methods to evaluate predictive validity can result in different findings and conclusions. Because predictive validity appears to be a crucial factor in the decision of which tool to use in practice (Bonta, 2002), the use of one performance indicator over another may influence adoption decisions. We review these three categories of performance indicators in greater detail below.

The first category of performance indicators measures whether assessments completed using a given instrument correctly identify groups of individuals who go on to commit an

antisocial act. Examples include the positive predictive value (PPV) and the number needed to detain (NND). These performance indicators typically are based on true positive and false positive information, though indices such as sensitivity also include false negative information (Altman & Bland, 1994). The second category of performance indicators measures whether assessments correctly identify groups of individuals who do not go on to commit an antisocial act. Examples include the negative predictive value (NPV) and the number safely discharged (NSD). These performance indicators typically are calculated using true negative and false negative information, though there are exceptions such as specificity, which includes false positive information (Altman & Bland, 1994). Acceptable false positive and false negative rates are context-specific (Smits, 2010); thus, benchmarks for interpreting these two categories of performance indicators have not been established in the risk assessment literature (Altman & Bland, 1994a, 1994b).

The third category of performance indicators provides global estimates of predictive validity by combining information on the frequency of true and false positives as well as true and false negatives (Glas, Lijmer, Prins, Bonsel, & Bossuyt, 2003). They are routinely reported with dispersion parameters such as standard errors or confidence intervals and either comparisons against chance estimates (p-values) or benchmarks to aid in interpretation (Ferguson, 2009; Rice & Harris, 2005). Examples of global performance indicators include the correlation coefficient ( $r$ ; the strength and direction of the association between risk classification and antisocial outcome), the odds ratio (OR; the ratio of the odds of an antisocial act in the high-risk group compared with the odds of an antisocial act in the low risk group), the hazard ratio (HR; the ratio of hazards at a single time for those who engaged in an antisocial act and those who did not), and the area under the curve (AUC; the probability that a randomly selected individual who committed an antisocial act received a higher risk classification than a randomly selected individual who did not).



The present review will focus on the use and reporting of the AUC. Briefly, the AUC is derived from the receiver operating characteristic (ROC) curve, a plot of true and false positive rates across a risk assessment tool's cut-off thresholds. The AUC has been recommended in the field of risk assessment to measure predictive validity due to its independence of cut-off thresholds and resistance to fluctuations in outcome base rates (Douglas, Otto, Desmarais, & Borum, 2012). However, use of the AUC as an indicator of predictive validity has also been criticized because it is difficult for non-specialists to understand (Munro, 2004), interpreted too optimistically (Sjöstedt & Grann, 2002), and unable to differentiate between instruments that produce assessments that accurately identify high- versus low-risk groups (Singh, Grann, & Fazel, 2011). Given these opposing viewpoints, the use and reporting of the AUC was of particular interest in this review.

### **Potential Influences on Analytic and Reporting Practices**

A variety of factors may influence the statistical methodologies used and performance indicators reported in predictive validity studies, such as the assessment approach. There are two general approaches to structured risk assessment: actuarial and structured professional judgment (SPJ). In the prediction-focused actuarial approach, weighted scores are assigned to criminal history, sociodemographic, and/or clinical factors empirically associated with the likelihood of antisocial behavior. These weighted scores are used to classify individuals into risk bins that correspond to probabilistic estimates of future antisocial behavior (Quinsey, Harris, Rice, & Cormier, 2006). Rather than prediction per se, SPJ instruments are intended to inform the development of individualized and coherent risk formulations as well as comprehensive risk management plans (Hart & Logan, 2011). As part of this process, the instruments act as aide-mémoires, guiding assessors to estimate risk across one of three final risk judgments (low, moderate, or high) after reviewing empirically- and theoretically-based risk and/or protective factors (Douglas, Ogloff, & Hart, 2003; Webster, Nicholls, Martin,

Desmarais, & Brink, 2006). Recent meta-analytic evidence suggests that actuarial and SPJ tools produce assessments with comparable predictive validity (Fazel, Singh, Doll, & Grann, 2012; Guy, 2008). However, how the predictive validity of actuarial and SPJ assessments is analyzed and reported has yet to be investigated. Given actuarial and SPJ instruments' divergent emphases on prediction versus prevention and risk management, respectively, it is possible that methodologies and performance indicators used to evaluate predictive validity differ as well.

Other factors that may influence predictive validity measurement practices include study authorship, geographic location, sample size, and date of publication. Just as predictive validity findings have been found to differ depending on whether an author of a risk assessment tool is also an author on a study investigating that tool (Blair, Marcus, & Boccaccini, 2008), analytic and reporting practices similarly may differ systematically as a function of authorship. Cross-cultural differences in researcher preferences or training additionally may be associated with use of certain methodologies and performance indicators, as may journal-specific requirements. Selection of a specific analytic approach may also be informed by consideration of statistical power, which is directly related to sample size. Finally, knowledge of different analytic approaches and acceptable reporting practices may have changed over time, making date of publication a potential source of variation.

### **The Present Review**

A second-order systematic review was conducted of published studies of risk assessment tools to investigate predictive validity measurement practices over the past 20 years. Specific aims included: (1) to identify the analytic approaches and performance indicators most commonly used to measure predictive validity; (2) to examine variability in how analytic approaches and performance indicators are defined and their results interpreted;

and (3) to explore whether the uses of analytic approaches and performance indicators differ as a function of the potential moderating factors outlined above.

### 2.3.3 Methods

#### **Study Selection**

Studies in the present review included 50 published articles from two systematic reviews on the predictive validity of structured risk assessments (Singh, Grann, & Fazel, 2011; Singh, Serper, et al., 2011). These two reviews focused on the comparative predictive validity of assessments produced by available instruments and did not examine methodological reporting practices for evidence of consistency. Thus, there is no overlap in the analyses and findings reported in those reviews and the present one.

The two reviews were selected for the breadth of their coverage of both actuarial and SPJ instruments; the former included predominantly studies of actuarial instruments and the latter predominantly studies of SPJ instruments. Specifically, the first review included studies published between January 1995 and November 2008 of risk assessment tools identified via surveys of clinicians to be the most commonly used in practice for assessing the risk of general violence, sexual violence, or criminal offending (Singh, Grann, & Fazel, 2011). The second review included studies published between January 1990 and January 2011 of tools designed for use in assessing community violence risk in psychiatric patients (Skeem & Monahan, 2011). Searches conducted for both reviews used acronyms and full names of tools as keywords in the PsycINFO, EMBASE, MEDLINE, and US National Criminal Justice Reference Service Abstracts databases. Studies in all languages from any country were considered for inclusion, and additional articles for both reviews were identified through reference lists, annotated bibliographies, and discussion with researchers in the field. Twenty-five studies were randomly selected from each review with duplicates excluded using the “runiform” command in STATA/ IC 10.1 for Windows (StataCorp., 2007). Three (6.0%)

**Table 12:** Characteristics of the 25 risk assessment tools included in the review

Type	Tool	Intended population	Intended outcome	Citation	
Actuarial	COVR	Civil psychiatric patients	Violent	Monahan et al. (2005)	
	GSIR	Adult offenders	Criminal	Nuffield (1982)	
	LS/CMI	Adult offenders	Criminal	Andrews et al. (2004)	
	LSI-R	Adult offenders	Criminal	Andrews and Bonta (1995)	
	MnSOST-R	Sexual offenders	Sexual	Epperson et al. (1998)	
	OGRS	Adult offenders	Criminal	Copas and Marshall (1998)	
	PCL-R	Non-specific (Adult)	N/A <sup>a</sup>	Hare (2003)	
	PCL:SV	Non-specific (Adult)	N/A <sup>a</sup>	Hart et al. (1995)	
	PCL:YV	Juvenile offenders	N/A <sup>a</sup>	Forth, Kosson, and Hare (2003)	
	RM2000	Sexual offenders	Violent + sexual	Thornton et al. (2003)	
	RRASOR	Sexual offenders	Sexual	Hanson (1997)	
	SORAG	Sexual offenders	Violent + sexual	Quinsey et al. (2006)	
	Static-99	Sexual offenders	Violent + sexual	Harris, Phenix, Hanson, and Thornton (2003)	
	Static-2002	Sexual offenders	Criminal + violent + sexual	Hanson and Thornton (2003)	
	VORAS	Adult offenders	Violent	Howells et al. (1997)	
	VRAG	Forensic psychiatric patients	Violent	Quinsey et al. (2006)	
	UK700	Civil psychiatric patients with psychosis	Violent	Wootton et al. (2008)	
	SPJ	HCR-20	Forensic psychiatric patients	Violent	Webster et al. (1997)
		HKT-30	Forensic psychiatric patients	Violent	Werkgroep (2002)
SORM		Forensic psychiatric patients	Violent	Grann et al. (2000)	
SARA		Spousal assaulters	Violent	Kropp et al. (1999)	
SAVRY		Adolescent offenders	Violent	Borum et al. (2003)	
SVR-20		Sexual offenders	Violent + sexual	Boer et al. (1997)	
V-RISK-10		Civil psychiatric patients	Violent	Hartvig et al. (2007)	
VRS		Forensic psychiatric patients	Violent	Wong and Gordon (2009)	

N/A, not applicable; Criminal, violent (including sexual) or non-violent; SPJ, structured professional judgment.

<sup>a</sup>Not designed to predict antisocial behavior.

studies were excluded because they did not use ROC curve methodology, an a priori focus of the present review. References for studies included in the review are provided in the Appendix.

The final sample of 47 studies examined the predictive validity of assessments produced using 25 instruments: the Classification of Violence Risk (COVR; Monahan et al., 2005;  $N_{\text{Studies}} = 4, 8.5\%$ ), the General Statistical Information on Recidivism (GSIR; Nuffield, 1982;  $N=1, 2.1\%$ ), the Historical, Clinical, Risk Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997; Webster, Eaves, Douglas, & Wintrup, 1995;  $N = 17, 36.2\%$ ), the Historische, Klinische, Toekomstige-30 (HKT-30; Werkgroep Pilotstudie Risicotaxatie Forensische Psychiatrie, 2002;  $N=2, 4.3\%$ ), the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004;  $N=1, 2.1\%$ ), the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995;  $N=2, 4.3\%$ ), the Minnesota Sex Offender Screening Tool-Revised (MnSOST-R; Epperson et al., 1998;  $N = 1, 2.1\%$ ), the Offender Group Reconviction Scale (OGRS; Copas & Marshall, 1998;  $N = 2, 4.3\%$ ), the Psychopathy Checklist-Revised (PCL-R; Hare, 1991, 2003;  $N=9, 19.1\%$ ), the Psychopathy Checklist: Screening Version (PCL:SV; Hart, Cox, & Hare, 1995;  $N = 8, 17.0\%$ ), the Psychopathy Checklist: Youth Version (PCL:YV; Forth, Kosson, & Hare, 2003;  $N = 2, 4.3\%$ ), the Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR; Hanson, 1997;  $N=2, 4.3\%$ ), the Risk Matrix 2000 (RM2000; Thornton et al., 2003;  $N = 1, 2.1\%$ ), the Spousal Assault Risk Assessment (SARA; Kropp, Hart, Webster, & Eaves, 1994, 1995, 1999;  $N = 1, 2.1\%$ ), the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2002, 2003;  $N = 4, 8.5\%$ ), the Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, & Cormier, 1998; Quinsey et al., 2006;  $N = 3, 6.4\%$ ), the Structured Outcome Assessment and Community Risk Monitoring (SORM; Grann et al., 2000;  $N=1, 2.1\%$ ), the Static-99 (Harris, Phenix, Hanson & Thornton, 1999; Hanson & Thornton, 2003;  $N=5,$

10.6%), the Static-2002 (Hanson & Thornton, 2003; N=3, 6.4%), the Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997; N=2, 4.3%), the Violence Risk Appraisal Guide (VRAG; Quinsey et al., 1998, 2006; N = 14, 29.8%), the Violence Risk Screening-10 (V-RISK-10; Hartvig et al., 2007; N = 1, 2.1%), the Violence Risk Scale (VRS; Wong & Gordon, 2009; N=1, 2.1%), the Violent Offender Risk Assessment Scale (VORAS; Howells, Watt, Hall, & Baldwin, 1997; N = 1, 2.1%), and an actuarial instrument developed as part of the UK700 study to predict violence in individuals diagnosed with psychotic disorders (Wootton et al., 2008; N=1, 2.1%). Three of the instruments – namely, the PCL-R, PCL:SV, and PCL:YV – were developed as personality assessments, but are commonly used in clinical practice for the purposes of violence risk assessment (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Khiroya, Weaver, & Maden, 2009; Viljoen, McLachlan, & Vincent, 2010). Twenty-four (51.1%) studies investigated more than one tool. Descriptive characteristics of included instruments are provided in Table 12.

The aim of the present review was not to describe the predictive validity achieved using the included instruments but rather to examine how predictive validity has been measured in studies investigating them. Readers interested in a comparison of the predictive validity of assessments produced by a subset of these instruments should consult the original reviews.

### **Data Extraction**

The first author extracted 37 study attributes and predictive validity reporting characteristics from the 47 included articles, with particular attention paid to descriptions and interpretations of ROC curve analysis and the AUC. As a measure of quality control, 10 (21.3%) studies were randomly selected and coded by a research assistant working independently of the authors. A high level of interrater agreement was established ( $k = 0.95$ ; Landis & Koch, 1977); disagreements were settled by consensus.

## Investigation of Heterogeneity

Chi-squared tests of differences in proportions (Altman, 1991) were used to investigate whether predictive validity measurement practices differed by the type of instrument under investigation (actuarial vs. SPJ), whether an author of the manual of an instrument under investigation was a study author, the geographic location in which the study was conducted (North America vs. Europe), the type of journal in which the article was published (general psychological or psychiatric audience vs. specialized audience), sample size ( $< 164$  vs.  $\geq 164$ ), and year of publication (before 2008 vs. 2008 and after).<sup>5</sup> A significance level of  $\alpha = 0.05$  was adopted for these analyses, which were conducted using MedCalc 11.3.8.0 for Windows (MedCalc Software, 2010).

### 2.3.4 Results

#### Study Characteristics

Study characteristics extracted from the 47 independent risk assessment studies are presented in Table 13. The majority of studies ( $N=43$ , 91.5%) were published in English and in a journal aimed at a specialized audience within psychology or psychiatry ( $N = 26$ , 55.3%). None of the 28 journals in which articles were published required the use of a particular analytic method or reporting of a particular performance indicator when investigating predictive validity. None of the articles followed a standardized reporting protocol. The titles or abstracts of most studies specified that predictive validity was measured ( $N = 44$ , 93.6%). Thirty-six (76.6%) studies examined an actuarial tool and 27 (57.4%) an SPJ tool, with 16 (34.0%) investigating both actuarial and SPJ instruments. An author of the manual for an instrument was also an author of an article investigating that instrument's predictive validity in 13 (27.7%) studies. Studies were conducted in 14 countries: the United Kingdom ( $N=14$ ;

---

<sup>5</sup> Median values were used to dichotomize sample size and year of publication. With respect to the latter, readers should note that the median year of publication may have been affected by the sampling strategy; specifically, one of the reviews from which we drew articles only included studies published through November 2008.

29.8%), Canada (N=8, 17.0%), the Netherlands (N=5, 10.6%), Sweden (N=5, 10.6%), the United States (N=4, 8.5%), Austria (N=2, 4.3%), Germany (N=2, 4.3%), Argentina (N=1, 2.1%), Belgium (N=1, 2.1%), Denmark (N=1, 2.1%), Finland (N=1, 2.1%), Norway (N=1, 2.1%), Serbia (N=1, 2.1%), and Spain (N = 1, 2.1%). The median sample size was 164 (interquartile range, IQR = 96–356, range=40–2681). A Cochran–Armitage  $w^2$  test of trend found that more predictive validity studies have been published in recent years,  $w^2(12, N = 47) = 13.43, p = 0.01$ .

### **Predictive Validity Measurement Characteristics**

Details regarding the analytic approaches and performance indicators used to measure predictive validity are provided in Table 13. In addition to ROC curve analysis, 21 (44.7%) studies employed at least one further methodology to measure predictive validity: 13 (27.7%) employed correlational analyses, nine (19.1%) logistic regression analyses, and three (6.4%) Cox survival analyses. In addition to the AUC, at least one further performance indicator was reported in 29 (61.7%) studies. In these 29 articles, additional estimates were most commonly those that indicated a tool's global predictive validity (i.e., the third category reviewed in the Introduction; N = 24, 82.8%) followed by those that indicated the ability to accurately identify groups of individuals most likely to commit an antisocial act (i.e., the first category; N = 10, 34.5%) and those that indicated the ability to accurately identify groups of individuals least likely to commit an antisocial act (i.e., the second category; N = 10, 34.5%). These additional performance indicators were defined in five (17.2%) and interpreted in eight (27.6%) of those studies in which they were reported.



**Table 13:** Reporting characteristics of studies investigating the predictive validity of 25 risk assessment tools

Category	Subcategory	Group	Actuarial	SPJ
			N of 36 (%)	N of 27 (%)
Study information	Type of journal	General	14 (38.9)	13 (48.1)
		Specific	22 (61.1)	14 (51.9)
	Geographical location	Europe	26 (72.2)	22 (81.5)
		North America	9 (25.0)	4 (14.8)
		South America	1 (2.8)	1 (3.7)
	Abstract claims study is a predictive validity study	Yes (with AUC as evidence)	13 (36.1)	10 (37.0)
		Yes (without AUC as evidence)	20 (55.6)	16 (59.3)
		No	3 (8.3)	1 (3.7)
	PI other than AUC reported	Yes	23 (63.9)	17 (63.0)
		No	13 (36.1)	10 (37.0)
ROC curve analysis	ROC curve accurately defined	Accurate definition	9 (25.0)	8 (29.6)
		Inaccurate definition	2 (5.6)	4 (14.8)
		No definition provided	25 (69.4)	15 (55.6)
	ROC curve provided as figure	Yes	11 (30.6)	7 (25.9)
		No	25 (69.4)	20 (74.1)
Area under the curve	AUC accurately defined	Accurate definition	10 (27.8)	5 (18.5)
		Inaccurate definition	6 (16.7)	4 (14.8)
		No definition provided	20 (55.6)	18 (66.7)
	AUC interpreted	Accurate interpretation	2 (5.6)	0 (0)
		Inaccurate interpretation	11 (30.6)	9 (33.3)
		Statement of comparability	14 (38.9)	15 (55.6)
		No interpretation provided	9 (25.0)	3 (11.1)

(Continues)

Category	Subcategory	Group	Actuarial	SPJ
			<i>N</i> of 36 (%)	<i>N</i> of 27 (%)
	Type of AUC calculated	Total score AUC	27 (75.0)	17 (63.0)
		Risk bin AUC	4 (11.1)	2 (7.4)
		Both total score and risk bin AUCs	2 (5.6)	8 (29.6)
		Unstated/Unclear	3 (8.3)	0 (0)
	<i>p</i> -values provided for AUC	Yes	25 (69.4)	18 (66.7)
		No (but term “significant” used)	4 (11.1)	3 (11.1)
		No (and term “significant” not used)	7 (19.4)	6 (22.2)
	SE or CI reported for AUC	Yes	25 (69.4)	23 (85.2)
		No	11 (30.6)	4 (14.8)
	PI reported when discussing previous literature <sup>a</sup>	AUC	20 (55.6)	12 (44.4)
		Correlation coefficient	4 (11.1)	4 (14.8)
		PPV	1 (2.8)	0 (0)
		None	13 (36.1)	15 (55.6)
		Yes	21 (58.3)	15 (55.6)
Risk bin schemes	Binning scheme acknowledged	No (but role of clinical judgment acknowledged)	N/A	2 (7.4)
		No	15 (41.7)	10 (37.0)
		Manual-suggested risk bins	6 (16.7)	9 (33.3)
	Risk bin outcome data provided	Alternative risk bins	15 (41.7)	6 (22.2)
		No risk bin data provided	15 (41.7)	12 (44.4)

SPJ, structured professional judgment; *N*, number of studies; PI, performance indicator; AUC, area under the curve; ROC, receiver operating characteristic; PPV, positive predictive value; SE, standard error; CI, confidence interval; N/A, not applicable.

<sup>a</sup>A single study could contribute multiple performance indicators.

*Receiver Operating Characteristic Curve*

The ROC curve was defined in 18 (38.3%) articles. Thirteen (72.2%) of these 18 studies defined the curve accurately as a plot of the true positive rate (sensitivity) against the false positive rate (1-specificity) for every possible cut-off threshold on an instrument. A figure depicting the ROC curve was provided in 12 (25.5%) of the included articles. When discussing ROC curve methodology, the three most commonly cited articles (in order) were those by Mossman (1994), Rice and Harris (1995), and Mossman and Somoza (1991).

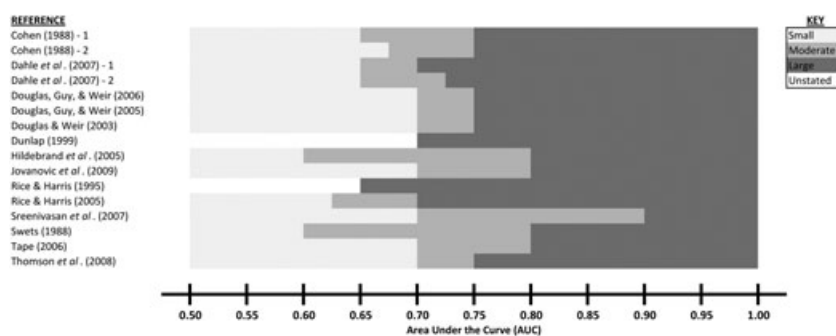
*Area Under the Curve*

Justifications for use of the AUC were provided in 26 (55.3%) studies. In these studies, justifications included base rate independence (N = 23, 88.5%), frequent use in the risk assessment literature (N=13, 50.0%), and lack of reliance on a cut-off threshold (N=11, 42.3%). The AUC was defined in 16 (34.0%) of all articles, with 10 (62.5%) of the 16 studies accurately defining the performance indicator as the probability that a randomly selected individual who committed an antisocial act will have received a higher risk classification than a randomly selected individual who did not. Incorrect definitions of the AUC included the proportion of individuals who committed an antisocial act who received higher risk scores than individuals who did not commit an antisocial act (N = 5, 31.3%), and the probability that a risk prediction will be accurate (N = 1, 6.3%). A dispersion parameter (standard error or confidence interval) was provided in 35 (74.5%) of all articles, and a p-value was reported in 32 (68.1%) studies.

Area under the curve values were interpreted in 16 (34.0%) of the 47 studies. In two (12.5%) of the 16 articles, authors accurately interpreted their AUC values with respect to the likelihood that a randomly selected individual who committed an antisocial act received a higher risk classification than a randomly selected individual who did not. In 14 (87.5%) of the 16 articles, AUC results were misinterpreted; specifically, AUCs were misinterpreted as

either the proportion of individuals whose outcome was correctly predicted by a risk assessment tool (N=8, 50.0%), or the proportion of individuals judged to be at high risk who went on to commit an antisocial act (i.e., the PPV; N = 6, 37.5%). Sixteen (34.0%) of the 47 included studies used benchmarks to determine whether AUCs were small, moderate, or large in magnitude; however, there was considerable variation in these benchmarks across studies, even when the same source was cited (see Figure 3). Finally, 21 (44.7%) of the included articles provided no interpretation of their AUC values, but stated that values were comparable with those reported in previous studies.

Limitations of the AUC were discussed in nine (19.1%) articles. In these nine articles, the most commonly reported limitation was the potential exaggeration of AUCs resulting from samples with low base rates of antisocial behavior, as small groups of high-scoring individuals who engage in antisocial acts result in large AUCs (N=4, 44.4%). Also noted was the insensitivity of AUCs to clinically relevant base rate information (N = 1, 11.1%), difficulty for non-experts to understand (N = 1, 11.1%), reliance on binary outcomes (N=1, 11.1%), inability to take time at risk into account (N = 1, 11.1%), lack of consideration as to the magnitude of score differences between individuals who committed antisocial acts and those who did not (N = 1, 11.1%), and routine evaluation of total risk score information rather than risk bins or final risk judgments (N = 1, 11.1%).



**Figure 3:** Variation in magnitude benchmarks for the area under the curve (AUC) in structured risk assessment tools.

*Risk Bins and Final Risk Judgments*

The majority of AUCs in both studies of actuarial (N=27 of 36, 75.0%) and SPJ (N = 17 of 27, 63.0%) tools were calculated using instruments' total scores rather than risk bins or final risk judgments, respectively. Study authors acknowledged that actuarial instruments were designed to estimate the probability of antisocial behavior in different risk bins using norms developed in calibration studies in 21 (58.3%) of the 36 actuarial articles, with six (16.7%) of the 36 articles reporting outcome information using manual-suggested bins. Two (5.6%) of the 36 actuarial instruments provided rates of antisocial behavior at each score. Fifteen (55.6%) of the 27 SPJ studies acknowledged that such instruments were designed to guide a professional's final judgment of low, moderate, or high risk. Outcome information for these three risk categories was provided in nine (33.3%) of the 27 SPJ articles.

**Investigation of Heterogeneity**

No significant differences in predictive validity reporting characteristics were found between studies of actuarial versus SPJ instruments ( $w^2[1, N=47] < 1.85, p > 0.18$ ), articles where a tool author was a study author compared with independent investigations ( $w^2[1, N=47] < 3.72, p > 0.05$ ), studies conducted in North America versus Europe ( $w^2[1, N=45] < 3.33, p > 0.07$ ), articles published in general versus specialized journals ( $w^2[1, N=47] < 1.98, p > 0.16$ ), analyses conducted on smaller versus larger samples ( $w^2[1, N=47] < 2.24, p > 0.13$ ), or articles published before compared with after 2008 ( $w^2[1, N = 47] < 2.66, p > 0.10$ ).<sup>6</sup>

---

<sup>6</sup> The largest  $\chi^2$  value and its associated p-value are reported for each set of tests of differences.

### 2.3.5 Discussion

Although several dozen systematic reviews have examined the predictive validity of risk assessment instruments, none has examined how this psychometric property has been measured (Singh & Fazel, 2010). To address this scientific gap, we conducted a second-order systematic review to investigate the analytic and reporting practices used in 47 studies concerning 25 risk assessment instruments. Published studies were identified from two recent systematic reviews and descriptively analyzed to identify those statistical methods and performance indicators most commonly used to investigate predictive validity. The consistency with which those methods and performance indicators were defined and interpreted was also explored, as were sources of between-study variability in measurement practices.

There were four principal findings of this review. First, the use of analytic methodologies (ROC curve analysis, correlational analysis, logistic regression, survival analysis) and performance indicators (AUC,  $r$ , OR, and HR) measuring a risk assessment instrument's global accuracy were much more common than those that measure the ability of an instrument to accurately identify groups of individuals at higher or lower risk of committing antisocial acts. In fact, the latter approaches were used in only a fifth of the articles. While the use of ROC curve analytic methodology was an inclusion criterion, only three of the 50 studies were excluded for this reason. These three studies used logistic regression and ORs to measure predictive validity.

Secondly, approximately two-thirds of the reviewed articles provided no definition of either the ROC curve or the AUC. Regarding interpretation, benchmarks for small, moderate, or large AUCs varied, even when the same source was cited. Performance indicators other than the AUC were defined in only about a tenth of articles and were interpreted in less than a fifth.

Thirdly, although virtually all of the included instruments were designed to either assign individuals to probabilistic risk bins or to aid in producing final risk judgments, fewer than half of the articles reported the predictive validity of such bins or judgments. When the predictive validity of risk bins or final risk judgments were examined, the bins or judgment categories recommended in the instruments' manuals were used in only a third of cases.

Finally, although measurement practices varied considerably across articles, they did not vary systematically as a function of risk assessment approach, study authorship, geographic location, type of journal, sample size, or year of publication. However, there were marginal trends ( $p < 0.10$ ) suggesting that studies conducted in North America included more accurate definitions of the AUC and more frequently provided ROC plots, and that the AUC was correctly defined more frequently in studies where a tool author was also a study author.

### **Implications**

The findings of the present review have implications for both researchers and practitioners. First, the lack of reporting consistency in the description and interpretation of performance indicators across studies suggests the need for standardized guidelines for risk assessment predictive validity studies. For example, the AUC was often incorrectly defined as the proportion of individuals who committed an antisocial act who received higher risk scores than individuals who did not commit an antisocial act, or the probability that a risk prediction will be accurate. In addition, the AUC was frequently misinterpreted, either as the proportion of the sample whose outcome was correctly predicted or as the proportion of the sample judged to be at high risk who went on to commit an antisocial act. To our knowledge, there currently exist no standardized reporting checklists for the prognostic risk assessment literature as there are for the medical diagnostic literature (e.g., the Standards for Reporting of Diagnostic Accuracy Studies statement; Bossuyt et al., 2003). Establishing such guidelines may improve the comparability of predictive validity reporting across studies and instruments.

Secondly, future studies may wish to report a variety of performance indicators to provide a comprehensive picture of the predictive abilities of assessments completed using a given tool. For example, a violence risk assessment instrument developed by Roaldset, Hartvig, and Bjørkly (2011) produced assessments with statistically significant AUCs, PPVs of 33% and NPVs of 98% at 3 months. If only global performance indicators had been reported, as was the case in four-fifths of studies in the present review, it would not have been known that assessments completed using the instrument were more accurate in identifying groups of individuals at low risk of future violence.

Thirdly, because AUC values representing small, moderate, or large magnitude effects varied from one study to the next, caution is warranted when using benchmarks to interpret ROC curve analysis findings. Decisions as to which risk assessment instrument to implement should not be based on this sole criterion, or, at least, on authors' interpretations of the AUC. Indeed, AUC values were misinterpreted in nine-tenths of studies in which an interpretation was offered.

Fourthly, in studies where total scores rather than actuarial risk bins or structured risk judgments are used to examine predictive validity, study authors should clarify that the validity of total scores and categorical estimates are not necessarily the same (Douglas et al., 2003; Meyers & Schmidt, 2008; Mills, Jones, & Kroner, 2005). Future research should continue exploring the extent to which the predictive validity of risk scores approximates that of categorical estimates. Regardless of the findings of such investigations, however, contingency information using manual-suggested actuarial risk bins or final risk judgments (i.e., the number of individuals assigned to each bin or judgment, and the number of those individuals who actually engaged in antisocial behavior) should be routinely reported in predictive validity studies. The current underreporting of such tabular data is an impediment to the use of more rigorous meta-analytic methodology (Singh, Grann, & Fazel, 2011).



## Limitations

The present review had several limitations. First, only published studies were included. While the exclusion of unpublished investigations is consistent with previous reviews of the risk assessment literature (Buchanan & Leese, 2001; Gerhold, Browne, & Beckett, 2007; Holtfreter & Cupp, 2007; Woods & Ashley, 2007), measurement practices may differ in these alternative dissemination formats. Second, only those studies that used ROC curve analysis were included in the review. However, only three studies were excluded to meet this criterion, suggesting that this is the methodology currently preferred by researchers for measuring the predictive validity of risk assessment tools and that this criterion should not have biased our findings. Third, the present investigation employed second-order systematic review methodology, using studies identified through the systematic searches of two recent reviews rather than conducting a new search and including all predictive validity studies therefrom. Relatedly, although the 25 tools investigated by the included studies represented the most commonly used actuarial and SPJ instruments according to recent surveys (Archer et al., 2006; Khiroya et al., 2009; Viljoen et al., 2010), over 125 other instruments have been developed to assess the risk of antisocial behavior in correctional and mental health populations (Singh, Serper, et al., 2011). Fourth, it is possible that the present review was underpowered to detect sources of between-study heterogeneity. Nonetheless, inconsistency in the description and interpretation of predictive validity findings appears to be a general phenomenon.

### 2.3.6 Conclusion

Research investigating the predictive validity of risk assessments produced using structured instruments has real world implications for tool selection and implementation. Because these instruments are used to inform important decisions related to public safety, the analytic approaches and performance indicators used to measure predictive validity warrant

consistent description and interpretation. The findings of the present review suggest that such consistency has yet to be achieved in the risk assessment literature. The construction of a standardized reporting quality checklist similar to those currently available in the diagnostic medical literature is currently underway and may both increase consistency and improve comparability of study findings. It is important to note, however, that the analytic approaches and performance indicators found to be used in the literature do not measure predictive validity vis-à-vis the likelihood that any one individual within a sample will engage in antisocial behavior (Hart, Michie, & Cooke, 2007). In future predictive validity studies, calculating a variety of performance indicators and routinely reporting outcome information for manual-suggested risk bins or professional judgments may prove helpful in clarifying the predictive abilities of risk assessments completed using structured instruments.

### 2.3.7 Acknowledgements

The authors thank Jonny Looms for his assistance with the interrater reliability check. The second author's (S.L.D.) work on this paper was partially supported by the National Institute on Drug Abuse (P30DA028807, PI: Roger H. Peters). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agency.

### 2.3.8 Appendix

#### **References for studies included in the review**

- Bengtson, S. (2008). Is newer better? A cross-validation of the Static-2002 and the Risk Matrix 2000 in a Danish sample of sexual offenders. *Psychology, Crime & Law, 14*, 85–106. DOI:10.1080/10683160701483104
- Canton, W. J., van der Veer, T. S., van Panhuis, P. J. A., Verheul, R., & van den Brink, W. (2004). De voorspellende waarde van risicotaxatie bij de rapportage pro Justitia: Onderzoek naar de HKT-30 en de klinische inschatting. [The predictive value of risk assessment in pro Justitia reports: A study of the HKT-30 and clinical assessment]. *Tijdschrift Psychiatrie, 46*, 525–535.

- Dahle, K. P. (2006). Strengths and limitations of actuarial prediction of criminal reoffense in a German prison sample: A comparative study of LSI-R, HCR-20 and PCL-R. *International Journal of Law & Psychiatry*, *29*, 431–442. DOI:10.1016/j.ijlp.2006.03.001
- 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. DOI:10.1002/cpp.452
- 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. DOI:10.1002/ab.20044
- Dolan, M., & 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. DOI:10.1037/1040-3590.20.1.35
- 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. DOI:10.1176/api.ps.54.10.1372
- 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 & Behavior*, *32*, 479–510. DOI:10.1177/0093854805278411
- Doyle, M., Carter, S., Shaw, J., & Dolan, M. (2011). Predicting community violence from patients discharged from acute mental health units in England. *Social Psychiatry & Psychiatric Epidemiology*, *189*, 520–526. DOI:10.1007/s00127-011-0366-8
- Doyle, M., & Dolan, M. (2006). Predicting community violence from patients discharged from mental health services. *British Journal of Psychiatry*, *189*, 520–526. DOI:10.1192/bjp.bp.105.021204
- Doyle, M., Shaw, J., Carter, S., & Dolan, M. (2010). Investigating the validity of the Classification of Violence Risk in a UK sample. *International Journal of Forensic Mental Health*, *9*, 316–323. DOI:10.1080/14999013.2010.527428
- Eher, R., Rettenberger, M., Schilling, F., & Pfafflin, F. (2008). Failure of Static-99 and SORAG to predict relevant reoffense categories in relevant sexual offender subtypes: A prospective study. *Sexual Offender Treatment*, *3*, 1–14.
- Gammelgård, M., Koivisto, A. M., Eronen, M., & Kaltiala-Heino, R. (2008). The predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) among institutionalised adolescents. *Journal of Forensic Psychiatry & Psychology*, *19*, 352–370. DOI:10.1080/14789940802114475
- Grann, M., Belfrage, H., & Tengström, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice & Behavior*, *27*, 97–114. DOI:10.1177/0093854800027001006
- Grann, M., Långström, N., Tengström, A., & Kullgren, G. (1999). Psychopathy (PCL-R) predicts violent recidivism among criminal offenders with personality disorders in Sweden.

*Law & Human Behavior*, 23, 205–217. DOI:10.1023/A:1022372902241

Grann, M., Sturidsson, K., Haggård-Grann, U., Hiscoke, U. L., Alm, P. O., Dernevik, M., et al. (2005). Methodological development: Structured Outcome Assessment and Community Risk Monitoring (SORM). *International Journal of Law & Psychiatry*, 28, 442–456. DOI:10.1016/j.ijlp.2004.03.011

Gray, N. S., Taylor, J., Fitzgerald, S., 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. DOI:10.1037/1040-3590.19.4.474

Gray, N. S., Taylor, J., & Snowden, R. J. (2011). Predicting violence using structured professional judgment in patients with different mental and behavioral disorders. *Psychiatric Research*, 187, 248–253. DOI:10.1016/j.psychres.2010.10.011

Gray, N. S., Taylor, J., & Snowden, R. J. (2008). Predicting violent reconvictions using the HCR-20. *British Journal of Psychiatry*, 192, 384–387. DOI:10.1192/bjp.bp.107.044065

Gray, N. S., Taylor, J., Snowden, R. J., MacCulloch, S., Phillips, H., & 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 & Clinical Psychology*, 72, 523–530. DOI:10.1037/0022-006X.72.3.523

Harris, G. T., Rice, M. E., & Camilleri, J. A. (2004). Applying a forensic actuarial assessment (the Violence Risk Appraisal Guide) to nonforensic patients. *Journal of Interpersonal Violence*, 19, 1063–1074. DOI:10.1177/0886260504268004

Helmus, L., & Hanson, R. K. (2007). Predictive validity of the Static-99 and Static-2002 for sex offenders on community supervision. *Sexual Offender Treatment*, 2, 1–14.

Jovanović, A. A., Tosevski, D. L., Ivković, M., Damjanović, A., & Gasić, J. (2009). Predicting violence in veterans with posttraumatic stress disorder. *Vojnosanitetski Pregled*, 66, 13–21.

Kelly, C. E., & Welsh, W. N. (2008). The predictive validity of the Level of Service Inventory-Revised for drug-involved offenders. *Criminal Justice & Behavior*, 35, 819–831. DOI:10.1177/0093854808316642

Kroner, C., Stadtland, C., Eidt, M., & Nedopil, N. (2007). The validity of the Violence Risk Appraisal Guide (VRAG) in predicting criminal recidivism. *Criminal Behaviour & Mental Health*, 17, 89–100. DOI:10.1002/cbm.644

Kropp, P. R., & Hart, S. D. (2000). The Spousal Assault Risk Assessment (SARA) guide: Reliability and validity in adult male offenders. *Law & Human Behavior*, 24, 101–118. DOI:10.1023/A:1005430904495

Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007). Actuarial assessment of risk for reoffense among adult sex offenders: Evaluating the predictive accuracy of the Static-2002 and five other instruments. *Criminal Justice &*

*Behavior*, 34, 37–59. DOI:10.1177/0093854806291157

- Lodewijks, H. P. B., Doreleijers, T. A. H., de Ruiter, C., & Borum, R. (2008). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during residential treatment. *International Journal of Law & Psychiatry*, 31, 263–271. DOI:10.1016/j.ijlp.2008.04.009
- Meyers, J. R., & Schmidt, F. (2008). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) with juvenile offenders. *Criminal Justice & Behavior*, 35, 344–355. DOI:10.1177/0093854807311972
- Mills, J. F., & Kroner, D. G. (2006). The effect of discordance among violence and general recidivism risk estimates on predictive accuracy. *Criminal Behaviour & Mental Health*, 16, 155–166. DOI:10.1002/cbm.623
- Monahan, J., Steadman, H. J., Robbins, P. C., Appelbaum, P., Banks, S., Grisso, T., et al. (2005). An actuarial model of violence risk assessment for persons with mental disorders. *Psychiatric Services*, 56, 810–815. DOI:10.1176/appi.ps.56.7.810
- Monahan, J., Steadman, H. J., Robbins, P. C., Silver, E., Appelbaum, P., Grisso, T., et al. (2000). Developing a clinically useful actuarial tool for assessing violence risk. *British Journal of Psychiatry*, 176, 312–319. DOI:10.1192/bjp.176.4.312
- Pham, T. H., Ducro, C., Marghem, B., & Reveillere, C. (2005). Evaluation du risque de recidive au sein d'une population de delinquants incarceres ou internes en Belgique francophone. [Prediction of recidivism among prison inmates and forensic patients in Belgium]. *Annales Medico Psychologiques*, 163, 842–845. DOI:10.1016/j.amp.2005.09.013
- Ramirez, M. P., Illescas, S. R., Garcia, M. M., Forero, C. G., & Pueyo, A. A. (2008). Predicción de riesgo de reincidencia en agresores sexuales. [Predicting risk of recidivism in sexual offenders]. *Psicothema*, 20, 205–210.
- Rettenberger, M., & Eher, R. (2007). Predicting reoffense in sexual offender subtypes: A prospective validation study of the German version of the Sexual Offender Risk Appraisal Guide (SORAG). *Sexual Offender Treatment*, 2, 1–12.
- Roadset, J., Hartvig, P., & Bjørkly, S. (2011). V-RISK-10: Validation of a screen for risk of violence after discharge from acute psychiatry. *European Psychiatry*, 26, 85–91. DOI:10.1016/j.eurpsy.2010.04.002
- Schaap, G., Lammers, S., & de Vogel, V. (2009). Risk assessment in female forensic psychiatric patients: A quasi-prospective study into the validity of the HCR-20 and PCL-R. *Journal of Forensic Psychiatry & Psychology*, 20, 354–365. DOI:10.1080/147899408002542873
- Sjöstedt, G., & Långström, N. (2002). Assessment of risk for criminal recidivism among rapists: A comparison of four different measures. *Psychology, Crime & Law*, 8, 25–40. DOI:10.1080/10683160208401807
- Snowden, R. J., Gray, N. S., & Taylor, J. (2010). Risk assessment for violence in individuals

- from an ethnic minority group. *International Journal of Forensic Mental Health*, 9, 118–123. DOI:10.1080/14999013.2010.501845
- Snowden, R. J., Gray, N. S., Taylor, J., & MacCulloch, M. J. (2007). Actuarial prediction of violent recidivism in mentally disordered offenders. *Psychological Medicine*, 37, 1539–1549. DOI:10.1017/S0033291707000876
- Sreenivasan, S., Garrick, T., Norris, R., Cusworth-Walker, S., Weinberger, L. E., Essres, G., et al. (2007). Predicting the likelihood of future sexual recidivism: Pilot study findings from a California sex offender risk project and cross-validation of the Static-99. *Journal of the American Academy of Psychiatry & the Law*, 35, 454–468.
- 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? *Legal & Criminological Psychology*, 4, 67–76. DOI:10.1348/135532599167798
- Sturup, J., Kristiansson, M., & Lindqvist, P. (2011). Violent behaviour by general psychiatric patients in Sweden: Validation of Classification of Violence Risk (COVR) software. *Psychiatry Research*, 188, 161–165. DOI:10.1016/j.psychres.2010.12.021
- Thomson, L., Davidson, M., Brett, C., Steele, J., & Darjee, R. (2008). Risk assessment in forensic patients with schizophrenia: The predictive validity of actuarial scales and symptom severity for offending and violence over 8–10 years. *International Journal of Forensic Mental Health*, 7, 173–189.
- van den Brink, R. H. S., Hooijschuur, A., van Os, T., Savenije, W., & Wiersma, D. (2010). Routine violence risk assessment in community forensic mental healthcare. *Behavioral Sciences & Law*, 28, 396–410. DOI:10.1002/bsl.904
- Wootton, L., Buchanan, A., Leese, M., Tyrer, P., Burns, T., Creedy, F., et al. (2008). Violence in psychosis: Estimating the predictive validity of readily accessible clinical information in a community sample. *Schizophrenia Research*, 101, 176–184. DOI:10.1016/j.schres.2007.12.490
- Wormith, J. S., Olver, M. E., Stevenson, H. E., & Girard, L. (2007). The long-term prediction of offender recidivism using diagnostic, personality, and risk/need approaches to offender assessment. *Psychological Services*, 4, 287–305. DOI:10.1037/1541-1559.4.4.287

- 2.4 **Singh, J. P.,** Yang, S., Mulvey, E., & the RAGEE Group (2014). Reporting standards for risk assessment predictive validity studies: The Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement. *Law & Human Behavior, 39*, 15-22.

#### 2.4.1 Abstract

Available reporting guidelines for prognostic and diagnostic accuracy studies apply primarily to biological assessment and outcomes, overlooking behavioral issues with major public health and safety implications such as violence. The present study aimed to develop the first set of reporting guidance for predictive validity studies of violence risk assessments: the Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement. A systematic search of 8 electronic databases prior to September 2012 identified 279 reporting guidelines for prognostic and diagnostic accuracy studies. Unique items were extracted and modified to make them relevant to risk assessment. A 4-wave Delphi process involving a multidisciplinary team of 37 international experts resulted in a 50-item reporting checklist. The panelists endorsed the RAGEE Statement checklist as being highly satisfactory and as indicating study features that should be reported routinely in manuscripts. Use of these proposed standards has the potential to improve the quality of the risk assessment literature.

#### 2.4.2 Introduction

Study quality has been shown to account for variation in clinical research findings (Rutjes et al., 2006). Because it is difficult to assess and compare study quality without transparent and consistent reporting of methodology, investigators in prognostic (McShane et al., 2006) and diagnostic medicine (Bossuyt et al., 2003) have developed well-received guidelines for methodological reporting in accuracy studies. Evidence suggests that the implementation of such guidelines has resulted in an improvement in reporting practices (Plint et al., 2006; Prady, Richmond, Morton, & MacPherson, 2008; Smidt et al., 2006; Smith

et al., 2008). However, available guidance of this type is limited to research assessing the risk of biological outcomes, overlooking behavioral issues with major public health and safety implications such as violence. Given the mortality rate and economic burden associated with violence, the World Health Organization (2002) has designated violence prevention as one of its priorities. This perspective is shared both by the mental health and criminal justice systems of numerous countries, including the United States and the United Kingdom, and is reflected in clinical guidelines for psychologists (American Psychological Association Presidential Task Force on Evidence-Based Practice, 2006), psychiatrists (American Psychiatric Association, 2004; National Institute for Health & Clinical Excellence, 2009), and nurses (Nursing & Midwifery Council, 2004) that recommend using evidence-based methods to assess violence risk. While the research base on the predictive validity of structured risk assessment instruments has grown exponentially (Buchanan, Binder, Norko, & Swartz, 2012), recent evidence from systematic reviews suggests that this literature has not achieved the same transparency and consistency as fields with established reporting guidelines. Considerable variability has been found in the reporting of essential sample- and study-level information in risk assessment studies published between 1990 and 2011 (Singh, Desmarais, & Van Dorn, 2013), making it difficult to assess the internal and external validity of their findings.

The development of reporting standards for violence risk assessment predictive validity studies could allow more informed comparisons between primary investigations, as well as sounder meta-analyses. This would, in turn, support the development of a cumulative science and potentially increase the reliability, utility, and impact of research in this area (Simera et al., 2010). Hence, to address the limitations of available reporting guidelines for prognostic and diagnostic accuracy studies when applied to the area of violence risk assessment, we used the Delphi technique to develop a novel reporting checklist: the Risk



Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement. Following published guidelines for developers of health research reporting guidance (Moher, Schulz, Simera, & Altman, 2010), our aim is to promote consistency and transparency for this important area of the behavioral sciences.

### 2.4.3 Method

#### **Design**

Consistent with the development of previous reporting standards (Hutchings, Raine, Sanderson, & Black, 2006), a multiwave Delphi process was used to select the item content of the RAGEE Statement. The Delphi method is based on the premise that group decisions are necessary when the scope of a problem is such that no single individual has sufficient expertise and knowledge to effect a solution. It is a structured communication technique that relies on the anonymous feedback of a panel of experts in an iterative process to establish consensus (Powell, 2003). By maintaining the anonymity of panelists and controlling their interactions, the Delphi technique avoids the disadvantages of more conventional consensus-based roundtable discussions and committees (Hasson, Keeney, & McKeena, 2000). Ethical review was waived by the University of South Florida Institutional Review Board; therefore, informed consent was not sought.

#### **Participants**

The Delphi panel consisted of 37 experts in the field of violence risk assessment (Table 14). This group included a multidisciplinary set of clinicians, researchers, legal professionals, and journal editors from 10 countries: Australia, Belgium, Canada, Germany, The Netherlands, Norway, Sweden, Switzerland, the United Kingdom, and the United States. The principal investigator (JPS) and coinvestigators (SY, EPM) organized, but were not members of, the Delphi panel. Potential panel members were identified by using recent reviews of the risk assessment literature (e.g., >; Hanson & Morton-Bourgon, 2009; Singh,

Serper, Reinharth, & Fazel, 2011; Skeem & Monahan, 2011) and were recruited to serve as experts if they met Farmer and Richman's (1965) criteria for Delphi panelist selection:

1. Extensive knowledge of the problem area and the ability to apply that knowledge
2. Good performance record in their particular area
3. High degree of objectivity and rationality
4. Time available to participate
5. Willingness to participate

## Materials

To identify a pool of items for consideration by the Delphi panel, a systematic search was performed to identify existing reporting guidance for prognostic and diagnostic accuracy studies. We searched the Cochrane Methodology Register, Database of Abstracts of Reviews of Effects, NHS Economic Evaluation Databases, Health Technology Assessment Databases, US National Criminal Justice Reference Service Abstracts, PROSPERO, PsycINFO, and MEDLINE prior to September 2012 using combinations of the following Boolean keywords: *prognos*<sup>□</sup>, *diagnos*<sup>□</sup>, *guid*<sup>□</sup>, *checklist*. Additional guidelines were identified using the EQUATOR (Enhancing the Quality and Transparency of Health Research) Network (Altman, Simera, Hoey, Moher, & Schulz, 2008), annotated bibliographies (e.g., Sanderson, Tatt, & Higgins, 2007), and discussion with experts.

Using this search strategy, we identified 279 published checklists (Figure 1). Items from each were extracted by the first and second authors with a high level of interrater agreement as established using a randomly selected subsample of 28 (10.0%) checklists ( $\kappa = 0.92$ ). Items addressing the same methodological principle (e.g., the inclusion of a structured abstract) were combined, and the wording of select items was modified to make them relevant to risk assessment (e.g., descriptions of biological tests were changed to descriptions of risk assessment instruments). This procedure, combined with a review of the literature on violence

(including sexual violence) and criminal recidivism risk assessment, resulted in the identification of 130 unique items.

### **Procedure**

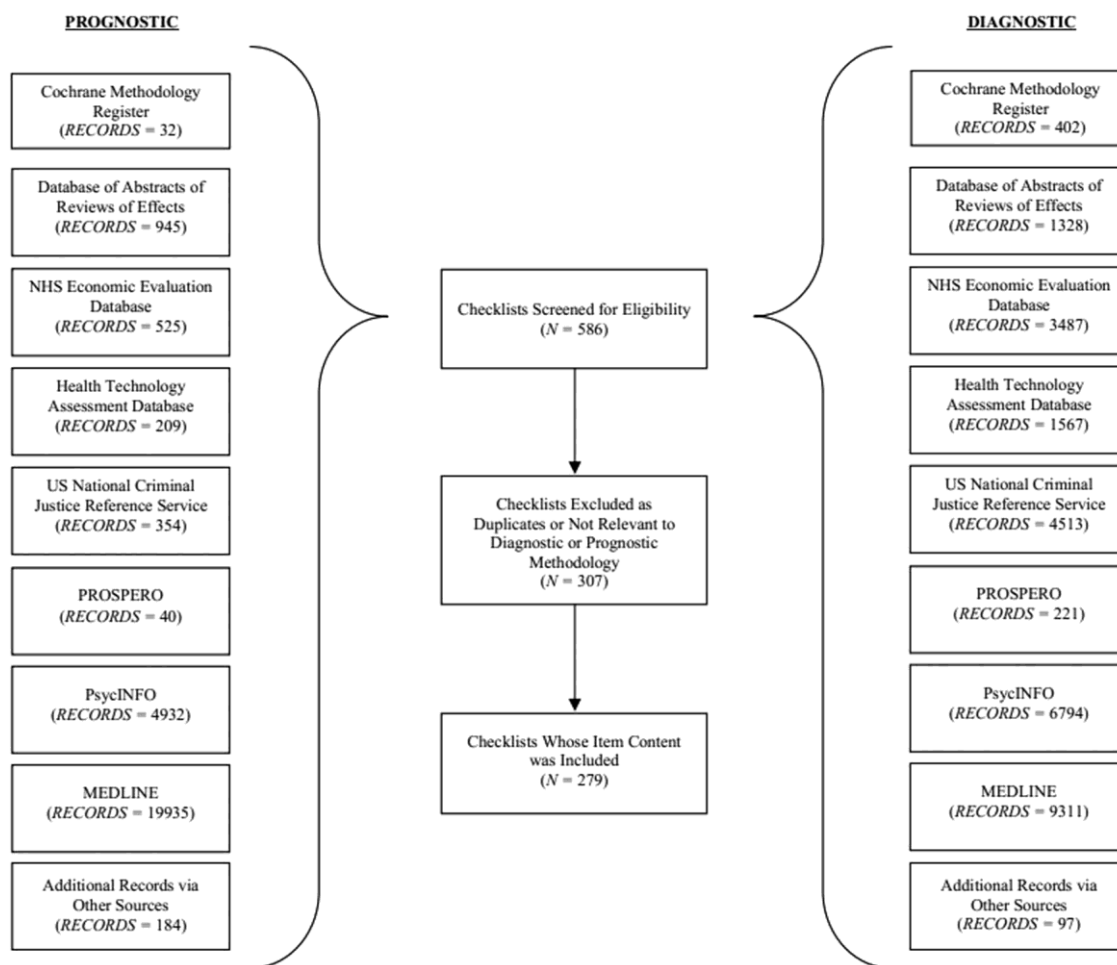
A four-wave Delphi process was conducted between September 2012 and February 2013 to select which of the 130 initially identified items would be included in the final RAGEE Statement. The Delphi process was conducted electronically using Qualtrics survey software ([www.Qualtrics.com](http://www.Qualtrics.com)), thus effectively managing the geographic dispersion of panelists and overcoming the time constraints related to physical meetings. Qualtrics has been used in recent research with forensic mental health professionals (e.g., Kimonis Fanniff, Borum, & Elliott, 2011; Singh, 2013) and has a number of benefits, including data collection through a secure server, libraries of customizable question templates, and a continuous file saving function to minimize data loss because of browser crashes.

In both the first and second waves of the Delphi process, panelists voted to definitely include, definitely exclude, or abstain from voting on each of the items. An inclusion threshold of 75% approval and an exclusion threshold of 25% disapproval were set (cf. Campbell, Piaggio, Elbourne, Altman, & the CONSORT Group, 2000). Items falling between these thresholds were retained for a further round of voting. Panelists had the opportunity to suggest new items, as well as to recommend modifications in wording. In the third wave, the panel was asked to dichotomously vote to either include or exclude remaining items. In the final wave, panelists used seven-item Likert scales to register their degree of satisfaction with the finished checklist (1 = *very dissatisfied*; 7 = *very satisfied*), as well as whether the guidance statement should be routinely used as reporting standards for risk assessment predictive validity studies (1 = *strongly disagree*; 7 = *strongly agree*). Upon the completion of each wave, approved items were summarized and panelists were given access to the voting results for each item if requested.

Response rates in each wave were maximized using the Dillman Total Design Method (Dillman, Smyth, & Christian, 2009). In accordance with this approach, an initial e-mail with an active Qualtrics link was sent to panelists on a Friday requesting participation in the i-ven wave. Three reminder e-mails were sent at seven day intervals after the initial distribution. Using this strategy, a 100% panelist response rate was achieved for each wave (Figure 4).

**Table 14:** Characteristics of the 25 risk assessment tools included in the review

Name	Affiliation at time of Publication
Stål Bjørkly, PsyD	Institute of Health Sciences, Molde University College, Molde, Norway
Marcus T. Boccacini, PhD	Department of Psychology and Philosophy, Sam Houston State University, Huntsville, Texas
Randy Borum, PsyD	School of Information, University of South Florida, Tampa, Florida
Alec Buchanan, PhD, MD	Department of Psychiatry, Yale University, New Haven, Connecticut
David J. Cooke, PhD	Psychology Department, Glasgow Caledonian University, Scotland, United Kingdom
Sarah L. Desmarais, PhD	Department of Psychology, North Carolina State University, Raleigh, North Carolina
Kevin S. Douglas, LLB, PhD	Department of Psychology, Simon Fraser University, Burnaby, BC, Canada
Michael Doyle, PhD, RMN	Institute of Brain, Behaviour and Mental Health, University of Manchester, Manchester, United Kingdom
John F. Edens, PhD	Department of Psychology, Texas A&M University, College Station, Texas
Eric B. Elbogen, PhD	Department of Psychiatry, University of North Carolina-Chapel Hill School of Medicine, Chapel Hill, North Carolina
Jerome Endrass, PhD	Department of Justice, Psychiatric/Psychological Service, Canton of Zurich, Switzerland
Seena Fazel, MD	Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, United Kingdom
Martin Grann, PhD, MBA	Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden
Laura S. Guy, PhD	Department of Psychiatry, University of Massachusetts Medical School, Worcester, Massachusetts
R. Karl Hanson, PhD	Public Safety Canada, Ottawa, ON, Canada
Robert D. Hare, PhD	Department of Psychology, University of British Columbia, Vancouver, BC, Canada
Grant T. Harris, PhD	Waypoint Centre for Mental Health Care Penetanguishene, ON, Canada
Stephen D. Hart, PhD	Department of Psychology, Simon Fraser University, Burnaby, BC, Canada
Kirk Heilbrun, PhD	Department of Psychology, Drexel University, Philadelphia, Pennsylvania
Mark A. Larsen, JD	Committee for Public Counsel Services, Commonwealth of Massachusetts, Boston, Massachusetts
John Monahan, PhD	School of Law, University of Virginia, Charlottesville, Virginia
Daniel F. Montaldi, PhD	Sexually Violent Predator Program, Florida Department of Children and Families, Tallahassee, Florida
Douglas Mossman, MD	Department of Psychiatry and Behavioral Neuroscience, University of Cincinnati College of Medicine, Cincinnati, Ohio
Tonia L. Nicholls, PhD	Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada
James R. P. Ogloff, JD, PhD	Centre for Forensic Behavioural Science, Monash University and Forensicare, Victoria, Australia
Randy K. Otto, PhD	Department of Mental Health Law & Policy, University of South Florida, Tampa, Florida
John Petrila, JD, LLM	Department of Health Policy and Management, University of South Florida, Tampa, Florida
Thierry H. Pham, PhD	Centre de Recherche en Defense Sociale, Tournai, Belgium
Martin Rettenberger, PhD	Department of Psychology, Johannes Gutenberg University Mainz, Mainz, Germany
Marnie Rice, PhD	Waypoint Centre for Mental Health Care Penetanguishene, ON, Canada
Corine de Ruiter, PhD	Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands
Astrid Rossegger, PhD	Department of Justice, Psychiatric/Psychological Service, Canton of Zurich, Switzerland
Nicholas Scurich, PhD	Department of Psychology & Social Behavior and Department of Criminology, Law & Society, University of California, Irvine, California
Jennifer L. Skeem, PhD	Department of Psychology & Social Behavior, University of California, Irvine, California
Robert L. Trestman, PhD, MD	Correctional Managed Health Care, University of Connecticut Health Center, Farmington, Connecticut
Frank Urbaniok, PhD	Department of Justice, Psychiatric/Psychological Service, Canton of Zürich, Switzerland
Jodi L. Viljoen, PhD	Department of Psychology, Simon Fraser University, Burnaby, BC, Canada



**Figure 4:** Identification of methodological reporting checklists for prognostic and diagnostic accuracy studies.

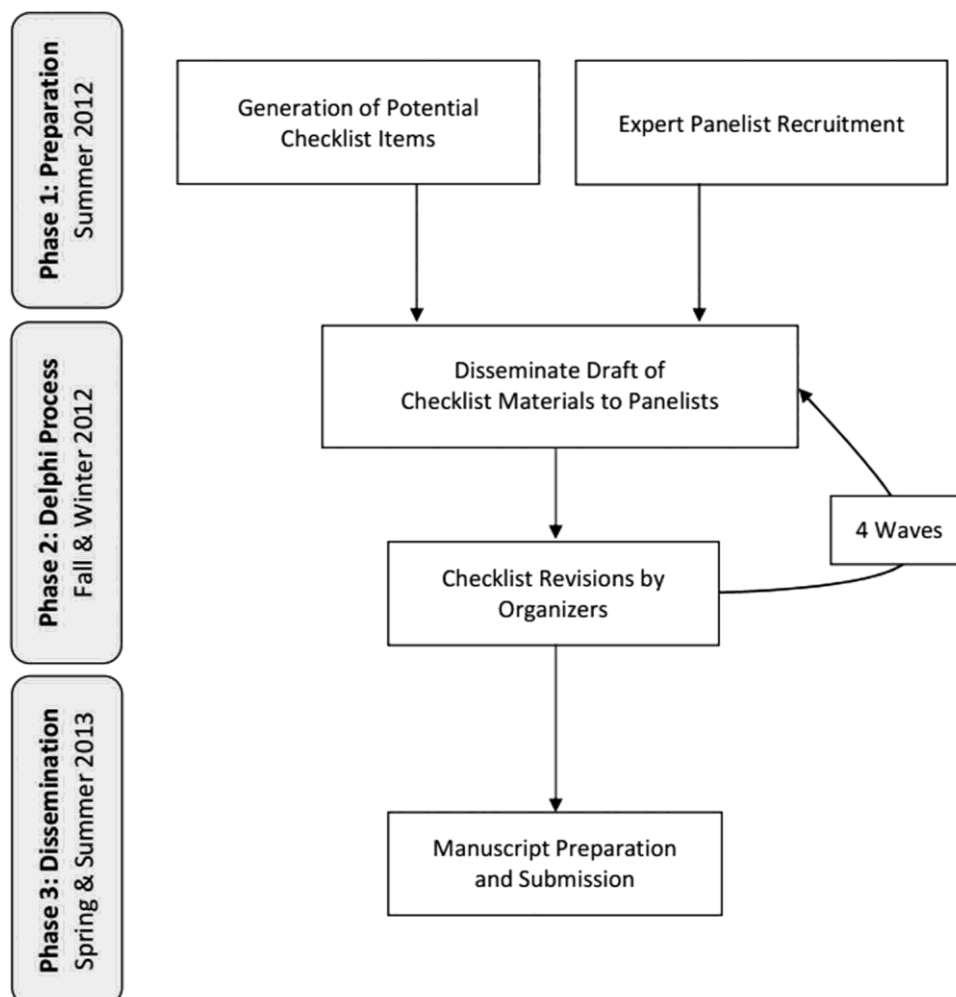
#### 2.4.4 Results

##### The RAGEE Statement Checklist Criteria

The completed RAGEE Statement includes 50 items and contains minimal reporting standards for the abstract ( $k$  items = 4), introduction ( $k = 2$ ), methods ( $k = 30$ ), results ( $k = 6$ ), and discussion ( $k = 4$ ) sections of risk assessment predictive validity manuscripts, as well as guidance on recommended disclosures ( $n = 4$ ; Table 15). The methods criteria are divided into six subsections: participants ( $k = 5$ ), instrument design ( $k = 7$ ), instrument administration ( $k = 5$ ), study design ( $k = 5$ ), predicted outcome ( $k = 2$ ), and statistical analysis ( $k = 6$ ). The

results criteria are divided into two subsections: participant outcomes ( $k = 2$ ) and predictive validity

( $k = 4$ ). The checklist version of the RAGEE Statement criteria can be found in the Supplemental Materials (Supplement 1). All criteria in the most comprehensive section of the checklist, concerning methods, can be met in fewer than 250 words (a sample methods section is available upon request), suggesting that the checklist does not place a substantial burden on authors. An elaboration document including exemplars for each item from the peer-reviewed predictive validity literature on risk assessment instruments was also developed to increase the usefulness of the checklist (Supplement 2).



**Figure 5:** Development of the Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) statement.

**Table 15:** Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement Checklist

Section	Item	Description	Endorsed (N of 37, %)
Abstract	1	Include a structured abstract describing the study	30 (81.1%)
	2	Identify the article as a risk assessment study in which predictive validity is measured	30 (81.1%)
	3	Identify the risk assessment instrument(s) whose predictive validity is measured	37 (100.0%)
	4	State the nature of the principal outcome (e.g., violence, sexual violence, criminal offending, institutional misconduct)	33 (89.2%)
Introduction	5	Provide the rationale and a summary of the scientific/theoretical background for the study	37 (100.0%)
	6	State the research questions and/or study aims	37 (100.0%)
Methods			
Participants	7	Report the sample size	37 (100.0%)
	8	Report the sex/gender composition of the sample	37 (100.0%)
	9	Report the average age at assessment (with dispersion parameter)	31 (83.8%)
Instrument design	10	Report the index offense composition of the sample	29 (78.4%)
	11	Report the characteristics of groups that underwent subgroup analysis	28 (75.7%)
	12	Report the acronym(s) and full name(s) of the instrument(s) under investigation with appropriate reference to source document	37 (100.0%)
	13	Report the number of items on the instrument(s) under investigation	30 (81.1%)
	14	Report the approach by which the assessment information from the instrument(s) under investigation is organized into an overall evaluation of risk	28 (75.7%)
	15	Report the population for which the instrument(s) under investigation was intended to be used	34 (91.9%)
	16	Report the outcome(s) that the instrument(s) under investigation was intended to assess	35 (94.6%)
	17	Report the length of follow-up for which manual-recommended probability estimates of risk were derived for the instrument(s) under investigation	30 (81.1%)
	18	Report the cut-off score(s) and/or risk categories that the instrument(s) under investigation was designed to use to classify risk level	30 (81.1%)
	Instrument administration	19	Report whether risk assessments were conducted in the context of research or practice
20		Identify when risk assessments occurred (e.g., pre-admission, admission, release, post-release)	37 (100.0%)
21		Report the number of assessors in the study as well as their training in the administration of the instrument(s) under investigation	34 (91.9%)
22		Identify the source(s) of information used to administer the instrument(s) under investigation	37 (100.0%)
Study design	23	Describe any modifications made to the instrument(s) under investigation	37 (100.0%)
	24	Report the geographical location and clinical setting in which risk was assessed	34 (91.9%)
	25	Describe the method(s) used to recruit participants	34 (91.9%)
	26	Identify the temporal design of the study (prospective or quasi-prospective)	36 (97.3%)
	27	Identify the setting in which participants were followed to ascertain whether the outcome(s) of interest had occurred	37 (100.0%)
	28	Report the average length of follow-up and time at risk (with dispersion parameter, if not fixed), including a description of periods subtracted from follow-up time (e.g., incarceration and/or hospitalization)	35 (94.6%)
Predicted outcome	29	Specify the event(s) coded as meeting outcome criteria (e.g., assault, rape, homicide)	34 (91.9%)
	30	Identify the type (e.g., arrest, charge, conviction, incarceration) and source (e.g., criminal records, self-report, collateral) of information used to detect outcome occurrence	37 (100.0%)
Statistical analysis	31	Describe the statistical methods used to conduct all analyses, and report the purpose of each analysis	30 (81.1%)
	32	Report whether risk scores and/or risk categories of the instrument(s) under investigation were used as an independent variable in analyses	32 (86.5%)
	33	Identify the statistical significance level used	34 (91.9%)
	34	Describe any subgroup analyses planned <i>a priori</i>	32 (86.5%)
	35	Report inter-rater reliability for administration of the instrument(s) under investigation (if conducted). If inter-rater reliability was not assessed, clarify why not	28 (75.7%)
	36	Include at least one discrimination performance indicator when measuring predictive validity	32 (86.5%)
Participant outcomes	37	Report the rate of attrition	32 (86.5%)
Predictive validity	38	Report the outcome occurrence rate for the entire sample as well as for relevant subgroups	34 (91.9%)
	39	Report predictive validity performance indicators for each outcome of interest as specified in the Methods with associated dispersion parameters	36 (97.3%)
	40	Report the number of participants with each risk score and/or in each risk category and how many went on to engage in the outcome(s) of interest	29 (78.4%)
	41	Report the results of subgroup analyses planned <i>a priori</i> as specified in the Methods	32 (86.5%)
	42	Describe and report the findings of any <i>post hoc</i> analyses conducted	28 (75.7%)
			(table continues)
Discussion	43	Provide a summary of the principal findings, including a discussion of their relevance in the context of the current literature	35 (94.6%)
	44	Discuss limitations of the study design	35 (94.6%)
	45	Discuss the generalizability of study findings	30 (81.1%)
	46	Discuss future research directions based on study findings	29 (78.4%)
Disclosures	47	Report any commercial interests and/or source(s) of funding as well as their role(s) in the conduct of the study	30 (81.1%)
	48	Report whether an author or translator of the risk assessment instrument(s) under investigation was also a study author	29 (78.4%)
	49	Report whether the study presented in the article has been published in an alternative form (e.g., government report)	28 (75.7%)
	50	Report whether the sample or a portion thereof has been studied in other publications	33 (89.2%)

### **Perceived Usefulness of the RAGEE Statement Checklist**

Using 7-point Likert scales, the average satisfaction rating with the checklist was 6.00 ( $SD = 1.04$ ), and the average support rating for using the checklist as reporting standards for risk assessment predictive validity studies was 5.84 ( $SD = 1.31$ ). Narrative comments revealed that lower ratings were due to the desire of some panelists to include mandatory reporting of calibration performance indicators (e.g., positive and negative predictive values) as an item rather than just discrimination performance indicators (e.g., area under the curve and correlation coefficients), the belief that no guidance should be given for introduction and discussion sections, and uncertainty about whether minimum reporting standards would exclude from consideration studies that merit publication.

#### 2.4.5 Discussion

The development of health research guidance has resulted in increased transparency and consistency in the methodological reporting of diagnostic and prognostic accuracy studies. None of this work, however, has been done in the critical and rapidly growing area of violence risk assessment. The creation of general guidelines for research studies such as the American Psychological Association *Journal Article Reporting Standards* (American Psychological Association Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008) has been a positive development for the social sciences, but such standards do not provide adequately specific guidance on sample and study-level characteristics that should be reported to maximize the clinical relevance of the risk assessment literature.

In the present report, we developed the first set of methodological reporting standards for predictive validity studies in violence risk assessment. A four-wave Delphi process involving 37 international experts from diverse fields resulted in a 50-item reporting checklist. Because the guidance statement was voted highly satisfactory and appropriate for



routine use as a reporting standard for risk assessment predictive validity studies, researchers may wish to reference the RAGEE Statement checklist while preparing manuscripts. In addition to being useful for manuscript authors, the use of the checklist by reviewers has the potential to expedite and increase agreement in the peer-review process.

Just as health research reporting guidance for other specialties has been adapted to related fields (Campbell, Piaggio, Elbourne, Altman, & the CONSORT Group, 2012; Ioannidis et al., 2004), the RAGEE Statement checklist may provide a useful basis for the development of methodological standards in other fields of behavioral prediction, such as suicide risk assessment. In its current form, however, the RAGEE Statement is designed for use only in studies of violence (including sexual violence) and criminal recidivism risk assessment.

Adherence to the RAGEE Statement guidance has the potential to resolve and overcome these obstacles to innovation, rigor, and relevance. It is important to note that the items on the RAGEE Statement checklist represent a *minimum* of what should be reported in risk assessment predictive validity studies at this time. Other valuable demographic, design, and performance information should continue to be reported where appropriate. For example, it is reasonable to assume that a brief summary of past predictive validity and reliability information will be reported in manuscripts. And as the field continues to develop, additional statistical approaches may enrich our picture of an instrument's predictive validity and expand the domain of study features that are desirable to report. It is our aim to update the RAGEE criteria as these developments arise. Hence, the RAGEE should be viewed as a living document. Meanwhile, when RAGEE Statement reporting criteria conflict with a journal's *Instructions for Authors*, please follow the latter.

#### 2.4.6 Conclusion

Mental health professionals are routinely called upon to assess the violence risk presented by their clients, frequently aided by structured instruments. Though a considerable literature exists on the predictive validity of these instruments, such studies are often plagued by inconsistent methodological reporting, limiting their reproducibility and clinical utility. The use of reporting guidelines has the potential to resolve and overcome these obstacles to innovation, rigor, and relevance.

- 2.5 Gmür, C., & Singh, J. P. (2014). Reporting quality of the sex offender risk assessment literature: A systematic review. *Sexual Offender Treatment, 9*, 1-14.

### 2.5.1 Abstract

The aim of the present study was to investigate the reporting quality of risk assessment predictive validity studies published in leading sex- and gender-related journals according to the Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement checklist. A systematic search for sex- and gender-related journals identified two that had published at least five articles investigating the predictive validity of a risk assessment method between 2008 and 2013: *Sexual Abuse: A Journal of Research and Treatment* (SAJRT; k articles = 21) and *Journal of Sexual Aggression* (JSA; k = 6). A RAGEE Statement checklist was coded for each article, and both sum score and item-level analyses were conducted. An average of 39.5 (SD = 3.0, range = 34-46) of the 50 RAGEE criteria were satisfied, suggesting that there is 20% room for improvement. SAJRT studies met significantly more RAGEE criteria than JSA studies. However, important sample- and study-level information needed to interpret findings was frequently missing across articles. Adopting RAGEE Statement guidelines into Instructions for Authors in journals that publish predictive validity studies on risk assessment methods for sex offenders has the potential to improve the quality and consistency of this literature.

### 2.5.2 Introduction

In recent years, the use of structured risk assessment tools to assess sex offender recidivism risk has been increasing in criminal justice systems, mental health services, and the interface between them. In Western countries including the United States, United Kingdom, Canada, Australia, and Switzerland such instruments are used to aide in decisions regarding

individual liberty and public protection (Helmus, Hanson, & Morton-Bourgon, 2011; Rossegger, Gerth, Urbaniok, Laubacher, & Endrass, 2010). Supporting the use of such instruments is a large evidence base on predictive validity that has grown exponentially in recent years (Singh, Fazel, Gueorguieva, & Buchanan, 2013).

A recent systematic review by Singh, Desmarais, and Van Dorn (2013) investigated the quality of methodological reporting in predictive validity studies of a number of structured instruments, including the following sex offender tools: Minnesota Sex Offender Screening Tool-Revised (Epperson et al., 1998), Rapid Risk Assessment for Sexual Offense Recidivism (Hanson, 1997), Risk Matrix 2000 (Thornton et al., 2003), Sex Offender Risk Appraisal Guide (Quinsey, Harris, Rice, & Cormier, 2006), Static-99 (Hanson & Thornton, 1999), Static-2002 (Hanson & Thornton, 2003), and Sexual Violence Risk-20 (Boer, Hart, Kropp, & Webster, 1997). The review concluded that "measurement practices varied considerably across articles" (p. 66), with key pieces of methodological information frequently going unreported (e.g., whether predictive validity estimates were based on total scores or categorical risk judgments). To address this, the first set of reporting guidance for risk assessment studies, the Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement, was recently published (Singh, Yang, Mulvey, & the RAGEE Group, in press). This 50-item reporting checklist was developed by a multidisciplinary panel of 37 international experts in the field of risk assessment from 10 countries. The panelists endorsed the RAGEE Statement checklist as being highly satisfactory and as indicating study features that should be reported routinely in risk assessment manuscripts. The innovative developments of the RAGEE Group appear promising, but in order to evaluate whether reporting practices are improving over time, it is first necessary to establish a baseline level of transparency in the current literature.

### **The Present Study**

The aim of the present review was to systematically assess the reporting quality of the sex offender risk assessment research literature in leading sex- and gender-related journals using the RAGEE Statement. Specifically, we aimed to establish a baseline estimate of the transparency achieved in the current literature, as well as to investigate whether such transparency fluctuates over time. Finally, we wished to examine differences in transparency by journal both descriptively as well as statistically. This review represents the first use of the RAGEE Statement to examine the transparency and consistency with which methodology and findings are reported in the sex offender risk assessment literature.

### 2.5.3 Method

#### **Systematic Search**

Sex- and gender-related journals were identified using a recently-published, comprehensive bibliography (Zucker, 2013) supplemented by consultation with experts in the field of sex offender risk assessment. Seventy-eight eligible journals were identified (see Supplement 1 for full list). The titles and abstracts of primary studies published between January 1, 2008 and November 1, 2013 were screened for evidence of predictive validity analyses having been conducted for either unstructured clinical judgments or use of a risk assessment tool to evaluate recidivism risk.<sup>1</sup> Journals were excluded if fewer than five articles including such predictive validity analyses had been published in the target timeframe, as reporting quality was not able to be compared using small-sample comparisons of exact levels (Larntz, 1978). Using this strategy, *Sexual Abuse: A Journal of Research and Treatment* (*SAJRT*;  $k$  articles = 21) and *Journal of Sexual Aggression* (*JSA*;  $k$  = 6) met inclusion criteria and underwent subsequent data extraction and analytic procedures. The included *SAJRT* articles included the following work: Brouillette-Alarie & Proulx, 2012; Craissati, Bierer, & South, 2010; Griffin, Beech, Print, Bradshaw, & Quayle, 2008; Griffin & Vettor, 2011;

Skelton & Vess, 2008; Wilcox, Beech, Markall, & Blacker, 2009. The JSA articles included the following work: Aebi, Plattner, Steinhausen, & Bessler, 2011; Barnett, Wakeling, & Howard, 2010; Beggs & Grace, 2010; Brown, Harkins, & Beech, 2012; Chu, Ng, Fong, & Teoh, 2012; Duwe & Freske, 2012; Eher, Matthes, Schilling, Haubner-MacLean, & Rettenberger, 2012; Fanniff & Letourneau, 2012; Grady, Edwards, Pettus-Davis, & Abramson, 2013; Grubin, 2011; Harkins, Beech, & Goodwill, 2010; Kingston, Yates, Firestone, Babchishin, & Bradford, 2008; McGrath, Lasher, & Cumming, 2012; Montana et al., 2012; Olver, Nicholaichuk, Gu, & Wong, 2013; Smallbone & Rallings, 2013; Storey, Watt, Jackson, & Hart, 2012; Swinburne Romine, Miner, Poulin, Dwyer, & Berg, 2012; Wakeling, Howard, & Barnett, 2011; Wilson, Abracen, Looman, Picheca, & Ferguson, 2011; Worling, Bookalam, & Litteljohn, 2012).

### **Data Extraction**

The first author coded a RAGEE Statement checklist for each of the 27 predictive validity articles. If a checklist criterion was satisfied, it was coded as a "1", and if it was not, as a "0". If a checklist criterion was not satisfied because it was not applicable to the sample, study design, or instrument under investigation, it was coded as a "1" for the purposes of this study. When more than one risk assessment method was investigated (e.g., the predictive validity of two instruments was compared) but a RAGEE criterion was only met for one of those methods (e.g., inter-rater reliability measured for one instrument but not another), then the criterion was coded as a "0" for the purposes of this study. When an article stated that methodological details were available in a previous manuscript, relevant RAGEE criteria were extracted from that previous manuscript.

As a measure of quality control, three (11.1%) of the included articles were randomly selected and coded by the corresponding author. A high level of inter-rater agreement was established ( $\kappa = 0.81$ ; Landis & Koch, 1977). Discrepancies were resolved upon discussion.

## Statistical Analyses

The proportion of articles that met each RAGEE criteria was descriptively examined both overall as well as for *SAJRT* and *JSA*, separately. In addition, Fisher's (1922) exact test, a statistical significance test measuring the association between two classification methods with small sample sizes, was used to investigate differences in proportions between the number of *SAJRT* articles that met each RAGEE criterion compared to *JSA* articles. Next, a Student's t-test was conducted to evaluate the difference in the overall number of RAGEE criteria met in *SAJRT* articles compared to *JSA* articles. Finally, a correlation coefficient (Spearman's  $\rho$ ) was calculated to test the presence and magnitude of a relationship between the number of RAGEE criteria satisfied and the date of publication. All analyses were conducted in MedCalc Version 11.3.8.0 for Windows using a standard significant threshold of  $\alpha = 0.05$ .

### 2.5.4 Results

#### Descriptive Characteristics

The average article fulfilled 39.5 ( $SD = 3.0$ , range = 34-46) of the 50 RAGEE criteria. An average of 2.7 ( $SD = 0.5$ , range = 1-3) of the four Abstract section criteria were met, 2.0 ( $SD = 0.0$ , range = 2-2) of the two Introduction section criteria were met, 22.0 ( $SD = 2.6$ , range = 17-27) of the 30 Method section criteria were met, 5.4 ( $SD = 0.6$ , range = 4-6) of the six Results section criteria were met, 3.5 ( $SD = 0.6$ , range = 2-4) of the four Discussion section criteria were met, and 3.9 ( $SD = 0.2$ , range = 3-4) of the four Disclosure section criteria were met. On average, the "Study Design" subsection of the Methods had the most criteria fulfilled ( $M = 4.0$  of 5 criteria,  $SD = 0.9$ , range = 2-5). The "Predicted Outcome" subsection of the Methods had the fewest criteria met ( $M = 1.3$  of 3 criteria,  $SD = 0.6$ , range = 0-2). The articles that satisfied the most RAGEE criteria included the works of Barnett and colleagues (2010;  $n$  criteria satisfied = 46, 92.0%), Faniff and colleagues (2012;  $n = 44$ , 88.0%), and Worling and colleagues (2012;  $n = 44$ , 88.0%).

### Association Findings

When differences in the proportion of articles meeting each RAGEE criterion were calculated between the two journals, Fisher's exact tests revealed that *SAJRT* articles were significantly more likely to: (1) identify the risk assessment instrument(s) whose predictive validity is measured in their Abstracts, (2) report the acronym(s) and full name(s) of the instrument(s) under investigation with appropriate referencing in their Methods, (3) report the population for which the instrument(s) under investigation was intended to be used in their Methods, (4) identify the statistical significance level used in their Methods, and (5) discuss the generalizability of study findings in their Discussions,  $\chi^2(1, N = 27) \geq 3.06, p < 0.05$ . *SAJRT* articles ( $M_n = 40.4, SD = 2.7$ ) met significantly more RAGEE criteria than *JSA* articles ( $M_n = 36.8, SD = 2.8$ ),  $t(25) = 2.86, p < 0.01$ . There was no association between date of publication and the total RAGEE score,  $\rho(25) = -0.03, p = 0.87$ .

### 2.5.5 Discussion

The present study represents the first systematic review of the reporting quality achieved in the current sex offender risk assessment literature using criteria set forth by the RAGEE Group. A systematic search identified two journals that had published at least five articles investigating the predictive validity of a risk assessment method between 2008 and 2013: *Sexual Abuse: A Journal of Research and Treatment* and *Journal of Sexual Aggression*. A RAGEE Statement checklist was coded for each article, and both sum score and item-level analyses were conducted. An average of 40 of the 50 RAGEE criteria were satisfied, implying an approximately 20% room for improvement. And this may be an overestimate, as it was assumed that not undisclosed competing interests did not exist, whereas recent systematic review evidence suggests that this may not always be the case (Singh, Grann, & Fazel, 2013). In addition, the majority of studies did not report whether modifications were made to



manual-based tool administration protocols, so it was assumed that no modifications were made. However, review evidence also exists that this is likely an overly conservative assumption (Rossegger et al., 2013). Overall, SAJRT articles were found to satisfy more RAGEE items than JSA, and no association was found between the date of publication and the number of RAGEE criteria satisfied, suggesting that the reporting quality of the sex offender risk assessment literature has remained stable over most of the past decade. It is important to note that the aim of this study (and the RAGEE initiative more generally) was not to investigate whether methodology and findings were reported correctly, but rather whether critical information in these areas was reported at all.

### **Implications**

The results of the present review have potentially important implications for researchers and practitioners. First, the finding that fundamental pieces of information on study design are frequently absent from published manuscripts supports the adoption of the RAGEE Statement as part of the *Instructions for Authors* for journals that publish research on sex offender risk assessment. This would serve as a quality control measure and is supported by leading researchers, practitioners, legal professionals, and journal editors. Second, readers of the current research literature on sex offender risk assessment may find it difficult to objectively assess the accuracy of study findings, as potentially important sample- and study-level information is frequently missing. For example, we found that how participants were recruited, their average age at assessment, when they were assessed, their index offense composition, and characteristics of the individuals who assessed them was not commonly reported. In addition, critical descriptive information on how many participants were classified into different risk levels, and how many of those persons went on to engage in the outcome of interest was reported in a minority of articles. Practitioners are advised that caution is warranted in interpreting the strength of reliability and validity research findings in

the absence of transparency (cf. Fava, 2007; Maj, 2005). Third, we have identified three articles that met the most RAGEE criteria that could be used as models for the predictive validity literature on sex offender risk assessment (Barnett et al., 2010; Faniff et al., 2012; Worling et al., 2012). Though these articles did not satisfy all 50 RAGEE conditions, they did fulfill over 85%.

### **Limitations and Future Directions**

The present review also has several limitations. First, articles published in journals not focused on issues of gender and sexuality did not meet inclusion criteria for the present review (e.g., *Law and Human Behavior*, *Journal of Interpersonal Violence*, *Criminal Justice and Behavior*). However, recent systematic reviews have found that predictive validity studies of sex offender risk assessment have been published in these more general forensic journals (Helmus, Hanson, Thornton, Babchishin, & Harris, 2012; Singh, Grann, & Fazel, 2011). Hence, future research on the quality of reporting in risk assessment studies should explore literature published in more general journals. Second, several eligible journals with predictive validity studies in the target timeframe were excluded (e.g., *Sexual Offender Treatment and Zeitschrift für Sexualforschung*), as fewer than five such articles were published between 2008 and 2013. This was necessary to conform to established statistical standards for the use of exact tests (Larntz, 1978), but further descriptive research may assist the esteemed editorial boards of these excluded journals in identifying areas of strength and potential areas for improvement in the reporting quality of their accepted manuscripts. Third, only peer-reviewed articles published between 2008 and 2013 were considered for inclusion. Albeit the aim of this review was to establish a baseline of reporting quality for the current literature on sex offender risk assessment, it should be kept in mind that influential research on this topic has been available for a number of decades (for a review see Hall, 1990). Future research may compare the reporting quality of articles on sex offender risk assessment across time to

examine trends. Given these limitations, our findings should be viewed as preliminary observations using a cross-section of the contemporary literature on sex offender risk assessment.

## **Conclusion**

Study quality has been shown to account for differences in research findings (Rutjes et al., 2006), and it is difficult to compare study quality without transparent and consistent reporting of methodology. The RAGEE Statement is the first set of expert reporting guidance set forth for the risk assessment literature to aid in this. The findings of the present study suggest that the adoption of standardized reporting guidelines such as the RAGEE Statement in journals that publish literature on sex offender risk assessment could benefit researchers, reviewers, and readers alike.

### **2.5.6 Conflict of Interest and Funding**

The authors are occasionally hired as experts for giving talks or workshops about risk assessment. Typically, this is done as part of the author's regular university duties (e.g. teaching students) but depending on the nature of the task and constituents, such activities are sometimes commissioned with remuneration.

The authors received no financial support for the research and/or authorship of this article.

**Table 16:** Reporting quality of two leading sex offender risk assessment journals according to Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement checklist criteria

RAGEE Checklist Section	Item	Description	SAJRT (k, %)	JSA (k, %)	chi <sup>2</sup>
<b>ABSTRACT</b>	R1	Include a structured abstract describing the study	0 (0)	0 (0)	
	R2	Identify the article as a risk assessment study in which predictive validity is measured	20 (95.2)	6 (100.0)	
	R3	Identify the risk assessment instrument(s) whose predictive validity is measured	21 (100.0)	4 (66.7) *	
	R4	State the nature of the principal outcome (e.g., violence, sexual violence, criminal offending, institutional misconduct)	17 (81.0)	6 (100.0)	
<b>INTRODUCTION</b>	R5	Provide the rationale and a summary of the scientific/theoretical background for the study	21 (100.0)	6 (100.0)	
	R6	State the research questions and/or study aims	21 (100.0)	6 (100.0)	
<b>METHODS</b>					
<i>Participants</i>	R7	Report the sample size	21 (100.0)	6 (100.0)	
	R8	Report the sex/gender composition of the sample	17 (81.0)	4 (66.7)	
	R9	Report the average age at assessment (with dispersion parameter)	1 (4.8)	0 (0)	
	R10	Report the index offense composition of the sample	7 (33.3)	2 (33.3)	
	R11	Report the characteristics of groups that underwent subgroup analysis	20 (95.2)	6 (100.0)	
<i>Instrument Design</i>	R12	Report the acronym(s) and full name(s) of the instrument(s) under investigation with appropriate reference to source document	21 (100.0)	4 (66.7) *	
	R13	Report the number of items on the instrument(s) under investigation	21 (100.0)	5 (83.3)	
	R14	Report the approach by which the assessment information from the instrument(s) under investigation is organized into an overall evaluation of risk	19 (90.5)	6 (100.0)	
	R15	Report the population for which the instrument(s) under investigation was intended to be used	21 (100.0)	4 (66.7) *	
	R16		9 (42.9)	0 (0)	

		Report the outcome(s) that the instrument(s) under investigation was intended to assess			
	R17	Report the length of follow-up for which manual-recommended probability estimates of risk were derived for the instrument(s) under investigation	8 (38.1)	3 (50.0)	
	R18	Report the cut-off score(s) and/or risk categories that the instrument(s) under investigation was designed to use to classify risk level	16 (76.2)	5 (83.3)	
<i>Instrument Administration</i>	R19	Report whether risk assessments were conducted in the context of research or practice	20 (95.2)	6 (100.0)	
	R20	Identify when risk assessments occurred (e.g., pre-admission, admission, release, post-release)	8 (38.1)	2 (33.3)	
	R21	Report the number of assessors in the study as well as their training in the administration of the instrument(s) under investigation	7 (33.3)	1 (16.7)	
	R22	Identify the source(s) of information used to administer the instrument(s) under investigation	18 (85.7)	6 (100.0)	
	R23	Describe any modifications made to the instrument(s) under investigation	21 (100.0)	6 (100.0)	
<i>Study Design</i>	R24	Report the geographical location and clinical setting in which risk was assessed	20 (95.2)	6 (100.0)	
	R25	Describe the method(s) used to recruit participants	11 (52.4)	4 (66.7)	
	R26	Identify the temporal design of the study (prospective or quasi-prospective)	14 (66.7)	5 (83.3)	
	R27	Identify the setting in which participants were followed to ascertain whether the outcome(s) of interest had occurred	19 (90.5)	3 (50.0)	
	R28	Report the average length of follow-up and time at risk (with dispersion parameter, if not fixed), including a description of periods subtracted from follow-up time (e.g., incarceration and/or hospitalization)	21 (100.0)	6 (100.0)	
<i>Predicted Outcome</i>	R29	Specify the event(s) coded as meeting outcome criteria (e.g., assault, rape, homicide)	9 (42.9)	0 (0)	
	R30	Identify the type (e.g., arrest, charge, conviction, incarceration) and source (e.g., criminal records, self-report, collateral) of information used to detect outcome occurrence	21 (100.0)	4 (66.7)	*

<i>Statistical Analysis</i>	R31	Describe the statistical methods used to conduct all analyses, and report the purpose of each analysis	18 (85.7)	4 (66.7)	
	R32	Report whether risk scores and/or risk categories of the instrument(s) under investigation were used as an independent variable in analyses	21 (100.0)	6 (100.0)	
	R33	Identify the statistical significance level used	15 (71.4)	1 (16.7)	*
	R34	Describe any subgroup analyses planned a priori	21 (100.0)	5 (83.3)	
	R35	Report inter-rater reliability for administration of the instrument(s) under investigation (if conducted). If inter-rater reliability was not assessed, clarify why not	10 (47.6)	1 (16.7)	
	R36	Include at least one discrimination performance indicator when measuring predictive validity	21 (100.0)	6 (100.0)	
<b>RESULTS</b>					
<i>Participant Outcomes</i>	R37	Report the rate of attrition	21 (100.0)	6 (100.0)	
	R38	Report the outcome occurrence rate for the entire sample as well as for relevant subgroups	20 (95.2)	5 (83.3)	
<i>Predictive Validity</i>	R39	Report predictive validity performance indicators for each outcome of interest as specified in the Methods with associated dispersion parameters	21 (100.0)	6 (100.0)	
	R40	Report the number of participants with each risk score and/or in each risk category and how many went on to engage in the outcome(s) of interest	9 (42.9)	3 (50.0)	
	R41	Report the results of subgroup analyses planned a priori as specified in the Methods	21 (100.0)	6 (100.0)	
	R42	Describe and report the findings of any post hoc analyses conducted	21 (100.0)	6 (100.0)	
<b>DISCUSSION</b>	R43	Provide a summary of the principal findings, including a discussion of their relevance in the context of the current literature	21 (100.0)	6 (100.0)	
	R44	Discuss limitations of the study design	17 (81.0)	6 (100.0)	
	R45	Discuss the generalizability of study findings	17 (81.0)	2 (33.3)	*
	R46	Discuss future research directions based on study findings	19 (90.5)	6 (100.0)	

DISCLOSURES				
R47	Report any commercial interests and/or source(s) of funding as well as their role(s) in the conduct of the study	21 (100.0)	5 (83.3)	
R48	Report whether an author or translator of the risk assessment instrument(s) under investigation was also a study author	21 (100.0)	6 (100.0)	
R49	Report whether the study presented in the article has been published in an alternative form (e.g., government report)	21 (100.0)	6 (100.0)	
R50	Report whether the sample or a portion thereof has been studied in other publications	21 (100.0)	6 (100.0)	

Note. \* $p < 0.05$ ;  $k$  = number of articles; SAJRT = *Sexual Abuse: A Journal of Research and Treatment* ( $k$  total = 21); JSA = *Journal of Sexual Aggression* ( $k$  total = 6);  $\chi^2$  = Fisher's exact test.

### 2.5.7 Supplement

#### Sex- and gender-related journals identified in systematic search process

- Archives of Sexual Behavior
- AIDS
- AIDS and Behavior
- AIDS Care
- AIDS Education and Prevention
- AIDS Patient Care and STDs
- AIDS Reviews
- Asian Journal of Andrology
- Asian Journal of Women's Studies • Australian Feminist Studies
- Body & Society
- Body Image
- Culture, Health, and Sexuality
- Differences: A Journal of Feminist Cultural Studies • Empirical Journal of Same-Sex Sexual Behavior
- European Journal of Women's Studies
- Evolution and Human Behavior
- Evolutionary Psychology
- Feminism and Psychology
- Feminist Review
- Feminist Studies
- Feminist Theory
- Feministische Studien
- Fertility and Sterility
- Frontiers: Journal of Women's Studies
- Gender and Education
- Gender and Society
- Gender, Place, and Culture
- GLQ: A Journal of Gay and Lesbian Studies
- Hormone Research in Pediatrics
- Hormones and Behavior

- Human Reproduction
- Indian Journal of Gender Studies
- International Journal of Andrology
- International Journal of Impotence Research
- International Journal of Sexual Health
- International Journal of STD and AIDS
- International Perspectives on Sexual and Reproductive Health
- Journal of Acquired Immune Deficiency Syndromes
- Journal of Andrology
- Journal of Child Abuse & the Law
- Journal of Child Sexual Abuse
- Journal of Gender Studies
- Journal of the History of Sexuality
- Journal of Homosexuality
- Journal of Men's Health
- Journal of Psychosomatic Obstetrics and Gynecology
- Journal of Sex & Marital Therapy
- Journal of Sex Research
- Journal of Sexual Aggression
- Journal of Sexual Medicine
- Journal of Sexual Offender Civil Commitment: Science and the Law
- Journal of Women's Health
- Men and Masculinities
- Menopause
- Paidika: The Journal of Paedophilia
- Perspectives on Sexual and Reproductive Health
- Politics and Gender
- Psychology of Men & Masculinity
- Psychology of Women Quarterly
- Psychoneuroendocrinology
- Sex Roles
- Sexual Abuse: A Journal of Research and Treatment
- Sexual and Relationship Therapy
- Sexual Development
- Sexual Health
- Sexual Offender Treatment
- Sexualities
- Sexuality and Disability
- Sexuality Research and Social Policy
- Sexually Transmitted Diseases
- Sexually Transmitted Infections
- Signs
- Women & Health
- Women and Therapy
- Women's Health Issues
- Women's Studies International Forum
- Zeitschrift für Sexualforschung



## RECORD OF ACHIEVEMENT

---

The research and the articles outlined in the present thesis were realized in collaboration with A number of colleagues. In the following, my independent contributions are listed per article.

### **Article 1**

Hurducas, C., Singh, J. P., de Ruiter, C., & Petrila, J. (2014). Violence risk assessment tools: A systematic review of clinical surveys. *International Journal of Forensic Mental Health, 13*, 181-192.

#### **My Contributions**

- Designed the systematic review protocol
- Developed data extraction sheet and supervised data extraction
- Participated in inter-rater reliability check
- Planned and supervised the statistical analysis
- Drafted and revised the manuscript

### **Article 2**

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., de Ruiter, C., Dean, K., Doyle, M., Folino, J., Godoy, V., Grann, M., Ho, R., Large, M., Nielsen, L. H., Pham, T., Rebocho, R. M., Reeves, K., Rettenberger, M., & Otto, R. K. (2014). Use and perceived utility of structured violence risk assessment tools in 44 countries: Findings from the IRiS Project. *International Journal of Forensic Mental Health, 13*, 193-206.

#### **My Contributions**

- Obtained competitive funding for the study
- Designed the study protocol
- Participated in developing the survey materials
- Participated in survey material dissemination
- Managed survey software
- Conducted the statistical analysis
- Drafted and revised the manuscript

**Article 3**

Singh, J. P., Desmarais, S., & Van Dorn, R. A. (2013). Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences & the Law, 31*, 55-73.

**My Contributions**

- Designed the second-order systematic review protocol
- Developed data extraction sheet and extracted data
- Participated in inter-rater reliability check
- Conducted the statistical analysis
- Drafted and revised the manuscript

**Article 4**

Singh, J. P., Yang, S., Mulvey, E., & the RAGEE Group (2014). Reporting standards for risk assessment predictive validity studies: The Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement. *Law & Human Behavior, 39*, 15-22.

**My Contributions**

- Designed the systematic review protocol
- Developed data extraction sheet and extracted data
- Participated in inter-rater reliability check
- Recruited and coordinated Delphi panel
- Managed survey software
- Conducted the statistical analysis
- Drafted and revised the manuscript

**Article 5**

Gmür, C., & Singh, J. P. (2014). Reporting quality of the sex offender risk assessment literature: A systematic review. *Sexual Offender Treatment, 9*, 1-14.

**My Contributions**

- Designed the systematic review protocol
- Developed data extraction sheet and supervised data extraction
- Participated in inter-rater reliability check
- Planned and supervised the statistical analysis
- Drafted and revised the manuscript

## REFERENCES

---

- Ægisdottir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2006). The meta-analysis of clinical judgement project: Fifty-six years of accumulated research on clinical versus statistical prediction. *Counseling Psychologist, 34*, 341–382. doi:10.1177/0011000005285875
- Abderhalden, C., Needham, I., Dassen, T., Halfens, R., Haug, H., & Fischer, J. (2006). Predicting inpatient violence using an extended version of the Brøset-Violence-Checklist: Instrument development and clinical application. *BMC Psychiatry, 6*(17), 1-9. doi: 10.1186/1471-244X-6-17
- Aebi, M., Plattner, B., Steinhausen, H.-C., & Bessler, C. (2011). Predicting sexual and nonsexual recidivism in a consecutive sample of juveniles convicted of sexual offences. *Sexual Abuse: A Journal of Research and Treatment, 23*(4), 456-473. doi: 10.1177/1079063210384634
- Almvik, R. (2008). *Assessing the risk of violence: Development and validation of the Brøset Violence Checklist* (Doctoral Dissertation). Norwegian University of Science and Technology, Trondheim, Norway.
- Almvik, R., & Woods, P. (1999). Predicting inpatient violence using the Broset Violence Checklist (BVC). *International Journal of Psychiatric Nursing Research, 4*(3), 498-505.
- Almvik, R., Woods, P., & Rasmussen, K. (2000). The Brøset Violence Checklist: Sensitivity, specificity, and interrater reliability. *Journal of Interpersonal Violence, 15*(12), 1284-1296.
- Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.

- Altman, D. G., & Bland, J. M. (1994a). Diagnostic tests 1: Sensitivity and specificity. *British Medical Journal*, *308*, 1552.
- Altman, D. G., & Bland, J. M. (1994b). Diagnostic tests 2: Predictive values. *British Medical Journal*, *309*, 102
- Altman, D. G., Simera, I., Hoey, J., Moher, D., & Schulz, K. (2008). EQUATOR: Reporting guidelines for health research. *Lancet*, *371*, 1149–1150. doi:10.1016/S0140-6736(08)60505-X
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders: Diagnostic criteria from DSM-IV*. Washington, DC: American Psychiatric Association.
- American Psychiatric Association. (2004). *Practice guidelines for the treatment of patients with schizophrenia*. Arlington, VA: Author.
- American Psychological Association Presidential Task Force on Evidence- Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*, 271–285. doi:10.1037/0003-066X.61.4.271
- American Psychological Association Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, *63*, 839 – 851. doi:10.1037/ 0003-066X.63.9.839
- Anderson, D., & Hanson, R.K. (2010).Static-99. In R.K. Otto & K.S. Douglas (Eds.), *Handbook of violence risk assessment* (251-267). New York, NY: Routledge Taylor & Francis Group.
- Andrews, D. A., & Bonta, J. (1995). *LSI-R: The Level of Service Inventory-Revised*. Toronto, ON: Multi-Health Systems.

- Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy & Law, 16*, 39–55. doi:10.1037/a0018362
- Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence, NJ: Matthew Bender & Company, Inc.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *LS/CMI: Level of Service/Case Management Inventory*. Toronto, ON: Multi-Health Systems.
- Archer, R., Buffington-Vollum, J.K., Stredney, R.V., & Handel, R.W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*, 84-94.
- Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110*, 486-498. doi: 10.1037/0033-2909.110.3.486
- Babbie, E. R. (2012). *The practice of social research*. Belmont, CA: Wadsworth Publishing Company.
- Barbaree, H.E., Langton, C.M., Blanchard, R., & Boer, D.P. (2008) Predicting recidivism in sex offenders using the SVR-20: The contribution of age-at-release. *International Journal of Forensic Mental Health, 7*, 47-64.
- Barnett, G. D., Wakeling, H. C., & Howard, P. D. (2010). An examination of the predictive validity of the Risk Matrix 2000 in England and Wales. *Sexual Abuse: A Journal of Research and Treatment, 22*(4), 443-470. doi: 10.1177/1079063210384274
- Bartel, P., Forth, A., & Borum, R. (2003). Development and concurrent validation of the Structured Assessment for Violence Risk in Youth (SAVRY). Unpublished manuscript.
- Beggs, S. M., & Grace, R. C. (2010). Assessment of dynamic risk factors: An independent validation study of the Violence Risk Scale: Sexual offender version. *Sexual Abuse: A Journal of Research and Treatment, 22*(2), 234-251. doi: 10.1177/1079063210369014

- Bengtson, S., & Pedersen, L. (2008). Implementation of evidence-based practices in forensic psychiatric clinical practice in Denmark: Are we there? *Scandinavian Journal of Forensic Sciences*, 2, 47–52.
- Bennett, C., Khangura, S., Brehaut, J. C., Graham, I. D., Moher, D., Potter, B. K., & Grimshaw, J. M. (2011). Reporting guidelines for survey research: An analysis of published guidance and reporting practices. *PloS Medicine*, 8, e1001069.
- Bjørkly, S., Hartvig, P., Roaldset, J. O., & Singh, J. P. (2014). Norwegian developments and perspectives on violence risk assessment. *Criminal Justice & Behavior*, 41(12), 1384–1397. doi: 10.1177/0093854814547949
- Blair, P. R., Marcus, D. K., & Boccaccini, M. T. (2008). Is there an allegiance effect for assessment instruments? Actuarial risk assessment as an exemplar. *Clinical Psychology: Science & Practice*, 15, 346–360. doi:10.1111/j.1468-2850.2008.00147.x
- 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.
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk-20. Professional guidelines for assessing risk of sexual violence*. Burnaby, BC: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice & Behavior*, 29, 355–379. doi:10.1177/0093854802029004002
- Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines, and training. *American Psychologist*, 9, 945–956. doi: 10.1037/0003-066X.51.9.945

- Borum, R., Bartel, P., & Forth, A. (2002). *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY)*. Tampa: University of South Florida.
- Borum, R., Bartel, P., & Forth, A. (2003). *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY). Version 1.1*. Tampa: University of South Florida.
- Borum, R., Bartel, P., & Forth, A. (2006). *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY)*. Odessa, FL: Psychological Assessment Resources.
- Borum, R., Lodewijks, H., Bartel, P. A., & Forth, A. E. (2010). Structured Assessment of Violence Risk in Youth (SAVRY). In R. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 63-79), New York, NY: Routledge/Taylor at Francis Group.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glaziou, P. P., Irwig, L. M., et al. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Clinical Chemistry*, *49*, 1–6. doi:10.1373/49.1.1
- Brouillette-Alarie, S., & Proulx, J. (2012). Predictive validity of the Static-99R and its dimensions. *Journal of Sexual Aggression*, *19*(3), 311-328. doi: 10.1080/13552600.2012.747630
- Brown, D. (2013, January 3). *Predicting violence is a work in progress*. The Washington Post. Retrieved from <http://www.washingtonpost.com>
- Brown, S., Harkins, L., & Beech, A. R. (2012). General and victim-specific empathy: Associations with actuarial risk, treatment outcome, and sexual recidivism. *Sexual Abuse: A Journal of Research and Treatment*, *24*(5), 411-430. doi: 10.1177/1079063211423944
- Buchanan, A., & Leese, M. (2001). Detention of people with dangerous severe personality disorders: A systematic review. *Lancet*, *358*, 1955–1959.

- Buchanan, A., Binder, R., Norko, M., & Swartz, M. (2012). Resource document on psychiatric violence risk assessment. *The American Journal of Psychiatry, 169*, 1–10. doi:10.1176/appi.ajp.2012.169.3.340
- Burns, K. E., Duffett, M., Kho, M. E., Meade, M. O., Adhikari, N. K., Sinuff, T., & Cook, D. J. (2008). A guide for the design and conduct of self-administered surveys of clinicians. *Canadian Medical Association Journal, 179*, 245–252. doi: 10.1503/cmaj.080372
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. M., & Kraemmer, B. (1989). *The Minnesota Multiphasic Personality Inventory-2 (MMPI-2) Manual for Administration and Scoring*. Minneapolis, MN: University of Minneapolis Press.
- Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *Minnesota Multiphasic Personality Inventory—Adolescent (MMPI-A): Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Caldwell, M. F. (2013). Accuracy of sexually violent person assessments of juveniles adjudicated for sexual offenses. *Sexual Abuse: A Journal of Research and Treatment, 25*, 516-526. doi:10.1177/1079063213480818
- Campbell, M. K., Piaggio, G., Elbourne, D. R., Altman, D. G., & the CONSORT Group. (2012). CONSORT 2010 statement: Extension to cluster randomised trials. *British Medical Journal, 345*, e5661. doi: 10.1136/bmj.e5661
- Campbell, S. M., Cantrill, J. A., & Roberts, D. (2000). Prescribing indicators for UK general practice: Delphi consultation study. *British Medical Journal, 321*, 425–428. doi:10.1136/bmj.321.7258.425



- Canton, W. J., Veer, T. S. van der, Panhuis, P. J. A. van, Verheul, R., & Brink, W. van den. (2004). *De betrouwbaarheid van risicotaxatie in de pro Justitia rapportage. Een onderzoek met behulp van de HKT-30* [The reliability of risk assessment in 'pro Justitia' reports. An investigation on the basis of the HKT-30]. *Tijdschrift voor Psychiatrie*, 46(8), 537-542.
- Catchpole, R., & Gretton, H. (2003). The predictive validity of risk assessment with violent young offenders: A 1-year examination of criminal outcome. *Criminal Justice & Behavior*, 30, 688-708.
- Chu, C. M., Daffern, M., Thomas, S., & Lim, J. Y. (2012). Violence risk and gang affiliation in youth offenders: A recidivism study. *Psychology, Crime & Law*, 18, 99–315. doi: 10.1080/1068316X.2010.481626
- Chu, C. M., Ng, K., Fong, J., & Teoh, J. (2012). Assessing youth who sexually offended: The predictive validity of the ERASOR, J-SOAP-II, and YLS/CMI in a non- Western context. *Sexual Abuse: A Journal of Research and Treatment*, 24(2), 153-174. doi: 10.1177/1079063211404250
- Cleckley, H. (1941). *The mask of sanity: An attempt to reinterpret the so-called psychopathic personality*. St. Louis MO: Mosby.
- Cobanoglu, C., Warde, B., & Moreo, P. J. (2001). A comparison of mail, fax and web-based survey methods. *International Journal of Market Research*, 43, 441–452.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46. doi: 10.1177/ 001316446002000104
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum.
- Conroy, M. A., & Murrie, D. C. (2007). *Forensic evaluation of violence risk: A guide to risk assessment and risk management*. New York: Oxford University Press.

- Cooper, B. S., Griesel, D., & Yuille, J. C. (2008). Clinical-forensic risk assessment: The past and current state of affairs. *Journal of Forensic Psychology Practice, 7*, 1-63. doi: 10.1300/J158v07n04\_01
- Copas, J., & Marshall, P. (1998). The Offender Group Reconviction Scale: A statistical reconviction score for use by probation officers. *Applied Statistics, 47*, 159–171. doi:10.1111/1467-9876.00104
- Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly, 64*, 464–494. doi:10.1086/318641
- 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 and Mental Health, 16*, 183-194.
- Craissati, J., Bierer, K., & South, R. (2010). Risk, reconviction and "sexually risky behaviour" in sex offenders. *Journal of Sexual Aggression, 17*(2), 153-165. doi: 10.1080/13552600.2010.490306
- Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys perceptions of burden. *Social Science Computer Review, 19*, 146–162. doi: 10.1177/089443930101900202
- Dahle, K. P., Schneider, V., & Ziethen, F. (2007). Standardisierte Instrumente zur Kriminalprognose [Standardized instruments for criminal prediction]. *Forensische Psychiatrie, Psychologie, Kriminologie, 1*, 15–26. doi:10.1007/s11757-006-0004-6
- Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993). DeMatteo, D., & Edens, J. F. (2006). The role and relevance of the Psy-chopathy Checklist-Revised in court: A case law survey of U.S. courts (1991-2004). *Psychology, Public Policy, and Law, 12*, 214–241. doi: 10.1037/1076-8971.12.2.214

- de Vries Robbé, M., de Vogel, V., Koster, K., & Bogaerts, S. (2014). Assessing protective factors for sexually violent offending with the SAPROF. *Sexual Abuse: A Journal of Research and Treatment*. doi 10.1177/1079063214550168
- de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbé, M. (2007). *Guide to the SAPROF. Structured Assessment of Protective Factors for Violence Risk. Version 1*. Utrecht: Forum Educatief.
- 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 & Human Behavior*, 28, 235-251.
- de Vries Robbé, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. I. (in press). Changes in dynamic risk and protective factors for violence during inpatient forensic psychiatric treatment: Predicting reductions in post discharge community recidivism. *Law & Human Behavior*.
- Dempster, R. J. (1998). *Prediction of sexually violent recidivism: A comparison of risk assessment instruments*. Unpublished Masters thesis. Department of Psychology, Simon Fraser University, Burnaby, BC, Canada.
- Department of Health. (2007). *Best practice in managing risk*. London: Department of Health.
- Desmarais, S. L., & Singh, J. P. (2013). *Instruments for assessing recidivism risk: A review of validation studies conducted in the U.S.* Lexington, KY: Council of State Governments.
- Dietiker, J., Dittmann, V., & Graf, M. (2007). Gutachterliche Risikoeinschätzung bei Sexualstraftätern Anwendbarkeit von PCL-SV, HCR-20+3 und SVR-20 [Risk assessment of sex offenders in a German-speaking sample: Applicability of PCL-SV, HCR-20+3, and SVR-20]. *Nervenarzt*, 78, 53-61.
- Dillman, D., Smyth, J., & Christian, L. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method (3rd ed.)*. Hoboken, NJ: John Wiley & Sons, Inc.

- Doctor, R., & Nettleton, S. (2003). *Dangerous patients: A psychodynamic approach to risk assessment and management*. London: Karnac.
- Dolan, M. C., & Rennie, C. E. (2008). The Structured Assessment of Violence Risk in Youth (SAVRY) as a predictor of recidivism in a U.K. cohort of adolescent offenders with conduct disorder. *Psychological Assessment, 20*, 35-46.
- Douglas, K. S., & Weir, J. (2003). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*.
- Douglas, K. S., Guy, L. S., & Weir, J. (2006). HCR-20 violence risk assessment scheme: Overview and annotated bibliography. Burnaby, BC: Simon Fraser University. Retrieved from <http://www.sfu.ca/psyc/faculty/hart/HCR-20%20Annotated%20Bibliograph,%202006.pdf>
- Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20<sup>V3</sup>: Assessing Risk for Violence: User guide*. Burnaby, BC: Mental Health, Law, and Policy Institute, Simon Fraser University. Retrieved from <http://www.sfu.ca/psychology/groups/faculty/hart/violink.htm>
- 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. doi:10.1176/api.ps.54.10.1372
- Douglas, K. S., Otto, R. K., Desmarais, S. L., & Borum, R. (2012). *Clinical forensic psychology*. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology, volume 2: Research methods in psychology* (pp. 213–244). Hoboken, NJ: John Wiley & Sons.
- Douglas, K. S., Shaffer, C., Blanchard, A. J. E., Guy, L. S., Reeves, K., & Weir, J. (2014). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*. Burnaby, BC: Mental Health, Law, and Policy Institute, Simon Fraser University.

- Drake, C., Clemente, L., & Perrin, G. (2006). *Violence risk assessment in clinical practice*. Phoenix, AZ: Arizona Psychologist. Retrieved from [http://azpsychologist.com/sourcebook\\_Violence\\_Risk-Lia%5B1%5D.doc](http://azpsychologist.com/sourcebook_Violence_Risk-Lia%5B1%5D.doc)
- Draugalis, J. R., Coons, S. J., & Plaza, C. M. (2008). Best practices for survey research reports: A synopsis for authors and reviewers. *American Journal of Pharmaceutical Education, 72*, 1–6.
- Dunlap, W. P. (1999). A program to compute McGraw and Wong's common language effect size indicator. *Behavior Research Methods, Instruments, & Computers, 31*, 706–709.
- Duwe, G., & Freske, P. J. (2012). Using logistic regression modeling to predict sexual recidivism: The Minnesota Sex Offender Screening Tool-3 (MnSOST-3). *Sexual Abuse: A Journal of Research and Treatment*. doi: 10.1177/1079063211429470
- Eher, R., Matthes, A., Schilling, F., Haubner-MacLean, T., & Rettenberger, M. (2012). Dynamic risk assessment in sexual offenders using STABLE-2000 and the STABLE-2007: An investigation of predictive and incremental validity. *Sexual Abuse: A Journal of Research and Treatment, 24*(1), 5-28. doi: 10.1177/1079063211403164
- Eifert, G. H., & Feldner, M. T. (2004). Conceptual foundations of behavioral assessment strategies: From theory to assessment. In M. Hersen (Ed.), *Comprehensive handbook of psychological assessment: Behavioral assessment* (pp. 95-107).
- Elbogen, E. B. (2002). The process of violence risk assessment: A review of descriptive research. *Aggression and Violent Behavior, 7*, 591–604. doi: 10.1016/S1359-1789(01)00051-9
- 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. doi:10.1037/1541-1559.2.2.133

- Elbogen, E. B., Mercado, C. C., Scalora, M. J., & Tomkins, A. J. (2002). Perceived relevance of factors for violence risk assessment: A survey of clinicians. *International Journal of Forensic Mental Health, 1*, 37–47. doi: 10.1080/14999013.2002.10471159
- Endrass, J., Danielsson, C., & Urbaniok, F. (2007). Assessing the therapeutic process with the forensic operationalized therapy/risk evaluation system (FOTRES). *European Psychiatry, 22*, S307-S308.
- Epperson, D. L., Kaul, J. D., Huot, S. J., Hesselton, D., Alexander, W., & Goldman, R. (1998). *Minnesota Sex Offender Screening Tool—Revised (MnSOST-R)*. St. Paul, MN: Minnesota Department of Corrections.
- Exner, J. E. (2003). *The Rorschach: A Comprehensive System. Vol. 1: Basic foundations and principles of interpretation (4th ed.)*. Hoboken, NJ: Wiley.
- Eysenbach, G. (2004). Improving the quality of web surveys: The checklist for reporting results of internet E-surveys (CHERRIES). *Journal of Medical Internet Research, 6*, e34. doi: 10.2196/jmir.6.3.e34
- Fanniff, A. M., & Letourneau, E. J. (2012). Another piece of the puzzle: Psychometric properties of the J-SOAP-II. *Sexual Abuse: A Journal of Research and Treatment, 24*(4), 378-408. doi: 10.1177/1079063211431842
- Farmer, R. N., & Richman, B. M. (1965). *Comparative management and economic progress*. Homewood, IL: Irwin.
- Fava, G. A. (2007). *Financial conflicts of interest in psychiatry*. *World Psychiatry, 6*, 19-24.
- Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). The prediction of violence and antisocial behaviour: A systematic review and meta-analysis of the utility of risk assessment instruments in 73 samples involving 24,827 individuals. *British Medical Journal, 345*, e4692.

- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology, Research and Practice, 40*, 532–538. doi:10.1037/a0015808
- Fisher, R. A. (1922). On the interpretation of  $\chi^2$  from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society, 85*, 87-94.  
doi:10.2307/2340521
- Foellmi, M., Rosenfeld, B., Rotter, M., Greenspan, M., & Khadivi, A. (2013, March). *Decisions, decisions. . . Violence risk triage practices in community psychiatric settings*. Presentation given at the American Psychology-Law Society Annual Conference, Portland, OR.
- Folino, J. O., & Castillo, J. L. (2006). Las facetas de la psicopatía según el Hare Psychopathy Checklist Revised y su confiabilidad. *Vertex, 69*, 325–330. Frye v. United States, 293 F. 1013 (1923).
- 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 & Comparative Criminology, 48*, 49–58.  
doi:10.1177/0306624x03257526
- Forth, A. E., Kosson, D. S., & Hare, R. D. (2003). *Hare Psychopathy Checklist-Revised: Youth Version*. Toronto: ON, Multi-Health Systems.
- Gerhold, C. K., Browne, K. D., & Beckett, R. (2007). Predicting recidivism in adolescent sexual offenders. *Aggression & Violent Behavior, 12*, 427–438.  
doi:10.1016/j.avb.2006.10.004
- Gilfoyle, N. F. P., Ogden, D. W., Tran, D. H., Friedman, S. S., Owens, A. L., & Pickering, W. C. (2011). *Brief for amici curiae American Psychological Association and Texas Psychological Association in support of petition for a writ of certiorari*. Washington, DC: American Psychological Association.

- Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., & Bossuyt, P. M. M. (2003). The diagnostic odds ratio: A single indicator of test performance. *Journal of Clinical Epidemiology*, *56*, 1129–1135. doi:10.1016/S0895-4356(03)00177-X
- Grady, M. D., Edwards, D., Pettus-Davis, C., & Abramson, J. (2013). Does volunteering for sex offender treatment matter? Using propensity score Analysis to understand the effects of volunteerism and treatment on recidivism. *Sexual Abuse: A Journal of Research and Treatment*, *25*(4), 319-346. doi: 10.1177/1079063212459085
- Grann, M., & Långström, N. (2007). Actuarial assessment of violence risk: To weigh or not to weigh? *Criminal Justice & Behavior*, *34*, 22-36.
- Grann, M., Haggård, U., Hiscoke, U. L., Sturidsson, K., Lovstrom, L., Siverson, E., et al. (2000). *The SORM manual*. Stockholm: Karolinska Institute.
- Green, B., Carroll, A., & Brett, A. (2010). Structured risk assessment in community forensic mental health practice. *Australasian Psychiatry*, *18*, 538–541. doi: 10.3109/10398562.2010.498513
- Griffin, H. L., & Vettor, S. (2011). Predicting sexual re-offending in a UK sample of adolescents with intellectual disabilities. *Journal of Sexual Aggression*, *18*(1), 64-80. doi: 10.1080/13552600.2011.617013
- Griffin, H. L., Beech, A., Print, B., Bradshaw, H., & Quayle, J. (2008). The development and initial testing of the AIM2 framework to assess risk and strengths in young people who sexually offend. *Journal of Sexual Aggression*, *14*(3), 211-225. doi: 10.1080/13552600802366593
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology*. Hoboken, NJ: John Wiley & Sons, Inc.
- Grubin, D. (2011). A large-scale evaluation of Risk Matrix 2000 in Scotland. *Sexual Abuse: A Journal of Research and Treatment*, *23*(4), 419-433. doi: 10.1177/1079063211411309



- Guy, L. (2008). *Performance indicators of the structured professional judgement approach for assessing risk for violence to others: A meta-analytic survey*. Unpublished doctoral dissertation, Simon Fraser University, Burnaby, BC.
- Guy, L. S., Douglas, K. S., & Hendry, M. C. (2010). The role of psychopathic personality disorder in violence risk assessments using the HCR-20. *Journal of Personality Disorders, 24*, 551–580. doi:10.1521/ pedi.2010.24.5.551
- Hall, G. C. N. (1990). Prediction of sexual aggression. *Clinical Psychology Review, 10*, 229-245. doi:10.1016/0272-7358(90)90059-J
- Hammond, S. M., & O'Rourke, M. M. (2000). *Developing a psychometric model for risk assessment: The case of the RAMAS*. Retrieved from: <http://ramas.co.uk/report2.pdf>
- Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism*. Ottawa, ON: Department of the Solicitor General of Canada.
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21*, 1–21. doi:10.1037/ a0014421
- Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders* (User Report 99–02). Ottawa, ON: Department of the Solicitor General of Canada.
- Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static-2002*. Ottawa, ON: Department of the Solicitor General of Canada.
- Hare, R. D. (1991). *Manual for the Revised Psychopathy Checklist* (1<sup>st</sup> ed.). Toronto, ON: Multi-Health Systems.
- Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised*. North Tonawanda, NY: Multi-Health Systems.

- Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised, 2nd Edition*. Toronto, ON: Multi-Health Systems.
- Harkins, L., Beech, A. R., & Goodwill, A. M. (2010). Examining the influence of denial, motivation, and risk on sexual recidivism. *Sexual Abuse: A Journal of Research and Treatment*, 22(1), 78-94. doi: 10.1177/1079063209358106
- Harris, G. T., & Rice, M. E. (2007). Characterizing the value of actuarial violence risk assessment. *Criminal Justice & Behavior*, 34, 297-313.
- Harris, G. T., Rice, M. E., Quinsey, V. L., & Cormier, C. A. (2015). *Violent offenders: Appraising and managing risk* (3<sup>rd</sup> ed.). Washington, DC: American Psychological Association.
- Harris, G. T., Rice, M. E., & Quinsey, V. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice & Behavior*, 20, 3–15. doi: 10.1177/ 0093854893020004001
- Hart, S. D., & Boer, D. P. (2010). Structured professional judgment guidelines for sexual violence risk assessment. In R.K. Otto & K.S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 269-294). New York, NY: Routledge Taylor & Francis Group.
- Hart, S. D., & Cooke, D. J. (2013). Another look at the imprecision of individual risk estimates made using actuarial risk assessment instruments. *Behavioral Sciences & the Law*, 31, 81–102. doi: 10.1002/bsl.2049
- Hart, S. D., & Logan, C. (2011). *Formulation of violence risk using evidence-based assessments: The structured professional judgment approach*. In P. Sturmey, & M. McMurrin (Eds.), *Forensic case formulation*. Chichester, UK: Wiley-Blackwell.
- Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *The Hare Psychopathy Checklist: Screening Version (PCL: SV)*. Toronto: Multi-Health Systems.

- Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: Evaluating the 'margins of error' of group v. individual predictions of violence. *British Journal of Psychiatry*, *49*, S60-S65. doi: 10.1192/bjp.190.5.s60
- Hart, S., Cox, D., & Hare, R. (1995). *The Hare Psychopathy Checklist: Screening Version*. Toronto, ON: Multi-Health Systems.
- Hartvig, P., Østberg, B., Alfarnes, S., Moger, T. A., Skjønberg, M., & Bjørkly, S. (2007). *Violence Risk Screening-10 (V-RISK-10)*. Oslo: Centre for Research and Education in Forensic Psychiatry.
- Hasson, F., Keeney, S., & McKeena, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*, 1008–1015. doi:10.1046/j.1365-2648.2000.t01-1-01567.x
- Hawley, C. J., Gale, T. M., Sivakumaran, T., & Littlechild, B. (2010). Risk assessment in mental health: Staff attitudes and an estimate of time cost. *Journal of Mental Health*, *19*, 88–98. doi:10.3109/09638230802523005
- Heilbrun, K. (2003). Violence risk: From prediction to management. In D. Carson & R. Bull (Eds.), *Handbook of Psychology in Legal Contexts* (2<sup>nd</sup> ed., pp. 127-143). Hoboken, NJ: Wiley.
- Heilbrun, K., Douglas, K. S., & Yasuhara, K. (2009). *Violence risk assessment: Core controversies*. In Skeem, J. L., Douglas, K. S., & Lilienfeld, S. O. (Eds.), *Psychological science in the courtroom: Consensus and controversy* (pp. 333–357). New York, NY: Guilford Press.
- Heilbrun, K., Dvoskin, J., Hart, S., & McNiel, D. (1999). Violence risk communication: Implications for research, policy, and practice. *Health, Risk & Society*, *1*, 91–106. doi:10.1080/13698579908407009

- Heilbrun, K., O'Neill, M., Stevens, T., Strohman, L., Bowman, Q., & Lo, Y. (2004).  
Assessing normative approaches to communicating violence risk: A national survey of  
psychologists. *Behavioral Sciences & the Law*, 22, 187–196. doi: 10.1002/bsl.570
- Helmus, L. (2008). *Annotated bibliography of Static-99 replications*. Retrieved from  
<http://www.static99.org/pdfdocs/static-99annotatedbibliography.pdf>
- Helmus, L., Babchishin, K. M., & Hanson, R. K. (2013). The predictive accuracy of the Risk  
Matrix 2000: A meta-analysis. *Sexual Offender Treatment*, 8(2), 1-20.
- Helmus, L., Hanson, K., & Morton-Bourgon, K. E. (2011). *International comparisons of the  
validity of actuarial risk tools for sexual offenders, with a focus on Static-99*. In Boer,  
D. P., Eher, R., Craig, L. A., Miner, M. H. & Pfäfflin, F. (Eds.), *International  
perspectives on the assessment and treatment of sexual offenders* (pp. 55-83): John  
Wiley & Sons, Ltd.
- Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., Harris, A. J. R. (2012). Absolute  
recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment  
tools vary across samples: A meta-analysis. *Criminal Justice & Behavior*, 33, 1148-  
1171. doi:10.1177/0093854812443648
- Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the  
predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age  
weights. *Sexual Abuse: A Journal of Research and Treatment*, 24, 64-101. doi:  
10.1177/1079063211409951
- Higgins, N., Watts, D., Bindman, J., Slade, M., & Thornicroft, G. (2005). Assessing violence  
risk in general adult psychiatry. *Psychiatric Bulletin*, 29, 131–133.  
doi:10.1192/pb.29.4.131
- Hildebrand, M., Hesper, B. L., Spreen, M., & Nijman, H. L. I. (2005). *De waarde van  
gestructureerde risico- taxatie en van de diagnose psychopathie: Een onderzoek naar*

- de betrouwbaarheid en predictieve validiteit van de HCR-20, HKT-30 en PCL-R. The value of structured risk assessment and the diagnosis psychopathy: Research into the reliability and predictive validity of the HCR-20, HKT-30 and the PCL-R.* Utrecht: Expertisecentrum Forensische Psychiatrie.
- Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *Counseling Psychologist, 34*, 400–409. doi: 10.1177/0011000005285877
- Ho, R. M. Y., Lau, J. S. F., Cheung, H. H. K., Lai, T. T. S., Tam, V. F. L., Chan, W. L., et al. (2013). Evaluation of a model of violence risk assessment (HCR-20) among adult patients discharged from a gazetted psychiatric hospital in Hong Kong. *Journal of Forensic Psychiatry & Psychology, 24*, 479–495. doi:10.1080/14789949.2013.809467
- Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly, 67*, 79–125. doi: 10.1086/346010?
- Holtfreter, K., & Cupp, R. (2007). Gender and risk assessment: The empirical status of the LSI-R for women. *Journal of Contemporary Criminal Justice, 23*, 363–382. doi:10.1177/1043986207309436
- Howells, K., Watt, B., Hall, G., & Baldwin, S. (1997). Developing programmes for violent offenders. *Legal & Criminological Psychology, 2*, 117–128. doi:10.1111/j.2044-8333.1997.tb00337.x
- Hox, J. J., & De Leeuw, E. D. (1994). A comparison of nonresponse in mail, telephone, and face-to-face surveys. *Quality and Quantity, 28*, 329–344. doi: 10.1007/BF01097014

- Hurducas, C., Singh, J. P., de Ruiter, C., & Petrila, J. (in this issue). Violence risk assessment tools: A systematic review of surveys. *International Journal of Forensic Mental Health*.
- Hutchings, A., Raine, R., Sanderson, C., & Black, N. (2006). A comparison of formal consensus methods used for developing clinical guidelines. *Journal of Health Services & Policy Research*, *11*, 218–224. doi: 10.1258/135581906778476553
- Inett, A., Wright, G., Roberts, L., & Sheeran, A. (2014). Predictive validity of the START with intellectual disabled offenders. *Journal of Forensic Practice*, *16*(1), 78-88. doi: 10.1108/JFP-12-2012-0029
- Interstate Commission for Adult Offender Supervision. (2007). *Sex offender assessment information survey*. ICAOS Documents 4-2007. Lexington, KY.
- Ioannidis, J. P. A., Evans, S. J. W., Gotzsche, P. C., O'Neill, R. T., Altman, D. G., Schulz, K., Moher, D., & the CONSORT Group. (2004). Better reporting of harms in randomized trials: An extension of the CONSORT Statement. *Annals of Internal Medicine*, *141*, 781–788. doi:10.7326/0003-4819-141-10-200411160-00009
- Jackson, R. L., & Hess, T. D. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sexual Abuse*, *19*, 425–448. doi: 10.1007/s11194-007-9062-3
- Jackson, R. L., Rogers, R., Neumann, C. S., & Lambert, P. L. (2002). Psychopathy in female offenders. An investigation of its underlying dimensions. *Criminal Justice & Behavior*, *29*, 692-704.
- Jackson, R.L., & Hess, D.T. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sexual Abuse: A Journal of Research and Treatment*, *19*, 409-448.
- Janus, E. (2004). Sexually violent predator laws: Psychiatry in service to a morally dubious enterprise. *Lancet*, *366*, 50–51. doi:10.1016/S0140-6736(04)17641-1

- Jovanović, A. A., Tosevski, D. L., Ivković, M., Damjanović, A., & Gasić, M. J. (2009). Predicting violence in veterans with posttraumatic stress disorder. *Vojnosanitetski Pregled*, *66*, 13–21.
- Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of Web and mail survey response rates. *Public Opinion Quarterly*, *68*, 94–101. doi:10.1093/poq/nfh006
- Kaser-Boyd, N. (2015). Threat assessment in homicide/suicide: The duty to warn. In: C. de Ruiter & N. Kaser-Boyd (2015). *Forensic psychological assessment in practice: Case studies*. New York, NY: Routledge.
- Keller, F., Kliemann, A., Karanediakova, D., Schnoor, K., Schuett, U., Keiper, P., ... & Schlaefke, D. (2011). Interrater reliability in the Forensic Operationalized Therapy/Risk Evaluation System. *Nervenheilkunde*, *30*(10), 813-817.
- Kelley, K., Clark, B., Brown, V., & Sitzia, J. (2003). Good practice in the conduct and reporting of survey research. *International Journal for Quality in Health Care*, *15*, 261–266. doi: 10.1093/intqhc/mzg031
- Khiroya, R., Weaver, T., & Maden, T. (2009). Use and perceived utility of structured violence risk assessments in English medium secure forensic units. *Psychiatrist*, *33*, 129–132. doi:10.1192/pb.bp.108.019810
- Kimonis, E. R., Fanniff, A., Borum, R., & Elliott, K. (2011). Clinician's perceptions of indicators of amenability to sex offender-specific treatment in juveniles. *Sexual Abuse: A Journal of Research & Treatment*, *23*, 193–211. doi:10.1177/1079063210384278
- Kingston, D. A., Yates, P. M., Firestone, P., Babchishin, K., & Bradford, J. M. (2008). Long-term predictive validity of the Risk Matrix 2000: A comparison with the Static-99 and the Sex Offender Risk Appraisal Guide. *Sexual Abuse: A Journal of Research and Treatment*, *20*(4), 466-484. doi: 10.1177/1079063208325206

- Kropp, P. R., Hart, S. D., & Belfrage, H. (2010). *The Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER), Version 2: User manual*. Vancouver, BC: ProActive ReSolutions.
- Kropp, P. R., Hart, S. D., Webster, C. W., & Eaves, D. (1994). *Manual for the Spousal Assault Risk Assessment guide*. Vancouver, BC: British Columbia Institute on Family Violence.
- Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1995) *Manual for the Spousal Assault Risk Assessment guide (2nd ed.)*. Vancouver, BC: British Columbia Institute on Family Violence.
- Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1999). *Spousal Assault Risk Assessment guide (SARA)*. Toronto, ON: Multi-Health Systems.
- Krysan, M., Schuman, H., Scott, L. J., & Beatty, P. (1994). Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly*, 58, 381–399.  
doi: 10.1086/269433
- Kwak, N., & Radler, B. (2002). A comparison between mail and web surveys: response pattern, respondent profile, and data quality. *Journal of Official Statistics*, 18, 257–273.
- Lally, S. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice*, 34, 491–498.  
doi:10.1037/0735-7028.34.5.491
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Large, M. M., Ryan, C. J., Singh, S. P., Paton, M. B., & Nielssen, O. B. (2011). The predictive value of risk categorization in schizophrenia. *Harvard Law Review*, 19, 25-33. doi: 10.3109/10673229.2011.549770



- Larntz, K. (1978). Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. *Journal of the American Statistical Association*, *73*, 253-263.  
doi:10.2307/2286650
- Lees-Haley, P. R., Smith, H. H., Williams, C. W., & Dunn, J. T. (1996). Forensic neuropsychological test usage: An empirical survey. *Archives of Clinical Neuropsychology*, *11*, 45-51. doi: 10.1016/0887-6177(95) 00011-9
- Leistico, A. M. R., Salekin, R. T., DeCoster, J., & Rogers, R. (2008). A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. *Law & Human Behavior*, *32*(1), 28-45.
- Lessard v. Schmidt, 349 F. Supp. 1078 (E.D. Wis. 1972).
- Linaker, O. M., & Busch-Iversen, H. (1995). Predictors of imminent violence in psychiatric inpatients. *Acta Psychiatrica Scandinavica*, *92*(4), 250-254. doi: 10.1111/j.1600-0447.1995.tb09578.x
- Lipsey, M. W., Chapman, G. L., & Landenberger, N. A. (2001). Cognitive-behavioral programs for offenders. *Annals of the American Academy of Political & Social Science*, *578*, 144-157.
- Lodewijks, H.P.B., Doreleijers, Th.A.H., de Ruiters, C., & Borum, R. (2008). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during residential treatment. *International Journal of Law & Psychiatry*, *31*, 263-271.
- Lofthouse, R. E., Lindsay, W. R., Totsika, V., Hastings, R. P., & Roberts, D. (2014). Dynamic risk and violence in individuals with an intellectual disability: Tool development and initial validation. *Journal of Forensic Psychiatry & Psychology*, *25*, 288-306.
- Logan, C. (2002). *Risk assessment among female inmates*. Speech presented to Towards a Safer Society. Violence: Origins Assessment and Management. Glasgow, Scotland.

- Maj, M. (2005). Conflicts of interests in psychiatric practice and research. *Psychiatrie*, 3, 138-140.
- Mason, T. (1998). Tarasoff liability: Its impact for working with patients who threaten others. *International Journal of Nursing Studies*, 35, 109-114.
- McEachran, A. (2001). *The predictive validity of the PCL:YV and the SAVRY in a population of adolescent offenders*. Unpublished master's thesis. Burnaby, British Columbia: Simon Fraser University
- McGrath, R. J., Lasher, M. P., & Cumming, G. F. (2012). The Sex Offender Treatment Intervention and Progress Scale (SOTIPS): Psychometric properties and incremental predictive validity with Static-99R. *Sexual Abuse: A Journal of Research and Treatment*, 24(5), 431-458. doi: 10.1177/1079063211432475
- McGrath, R.J., Cumming, G.F., & Burchard, B.L. (2003). *Current practices and trends in sexual abuser management: The Safer Society 2002 Nationwide Survey*. Brandon, VT: Safer Society Foundation, Inc.
- McShane, L. M., Altman, D. G., Sauerbrei, W., Taube, S. E., Gion, M., Clark, G. M. (2006). Reporting recommendations for tumor Marker prognostic studies (REMARK). *Breast Cancer Research & Treatment*, 100, 229–235. doi:10.1007/s10549-006-9242-8
- McSherry, B., & Keyzer, P. (2009) *Sex offenders and preventive detention: Politics, policy and practice*. Annandale, VA: The Federation Press.
- Medin, C., Roy, S. & Ann, T. (1999). *World Wide Web versus mail surveys: A comparison and report*. Paper presentation at ANZMAC99 Conference, Marketing in the Third Millennium, Sydney, Australia.
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and review of the evidence*. Minneapolis: University of Minnesota Press.

- Mental Health Centre Penetanguishene. (2009). *Replications of the Violence Risk Appraisal Guide or Sex Offender Risk Appraisal Guide in assessing violence risk*. Retrieved March 9, 2010, from [http://www.mhcp.on.ca/Site\\_Published/internet/-SiteContent.aspx](http://www.mhcp.on.ca/Site_Published/internet/-SiteContent.aspx)
- Meyers, J. R., & Schmidt, F. (2008). Predictive validity of the Structured Assessment for Violence Risk in Youth (SAVRY) with Juvenile Offenders. *Criminal Justice & Behavior, 35*, 344–355. doi:10.1177/93854807311972
- Meyers, J., & Schmidt, F. (2008). Predictive validity of the Structured Assessment for Violence Risk in Youth (SAVRY) with juvenile offenders. *Criminal Justice & Behavior, 35*, 344-355.
- Mills, J. F., Jones, M. N., & Kroner, D. G. (2005). An examination of the generalizability of the LSI-R and VRAG probability bins. *Criminal Justice & Behavior, 32*, 565–585. doi:10.1177/0093854805278417
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*, e1000097.
- Moher, D., Schulz, K. F., Simera, I., & Altman, D. G. (2010). Guidance for developers of health research reporting guidelines. *PLoS Medicine, 7*, e10000217. doi:10.1371/journal.pmed.1000217
- Monahan, J. (1981). *The clinical prediction of violent behavior*. Rockville, MD: U.S. Department of Health and Human Services.
- Monahan, J. (2006). A jurisprudence of risk assessment: Forecasting harm among prisoners, predators and patients. *Virginia Law Review, 92*, 391–435.
- Monahan, J. (2006). *Tarasoff* at thirty: How developments in science and policy shape the common law. *University of Cincinnati Law Review, 75*, 497-521.

- Monahan, J., Steadman, H., Appelbaum, P., Grisso, T., Mulvey, E., Roth, L., ... & Silver, E. (2005). *Classification of Violence Risk*. Lutz, FL: Psychological Assessment Resources.
- Monahan, J., Steadman, H., Silver, E., Appelbaum, P., Robbins, P., Mulvey, E., ... & Banks, S. (2001). *Rethinking risk assessment: The MacArthur Study of Mental Disorder and Violence*. New York: Oxford University Press.
- Montana, S., Thompson, G., Ellsworth, P., Lagan, H., Helmus, L., & Rhoades, C. J. (2012). Predicting relapse for catholic clergy sex offenders: The use of the Static-99. *Sexual Abuse: A Journal of Research and Treatment*, 24(6), 575-590. doi: 10.1177/1079063212445570
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting & Clinical Psychology*, 62, 783–792.
- Mossman, D., & Somoza, E. (1991). ROC curves, test accuracy, and the description of diagnostic tests. *Diagnostic Testing in Neuropsychiatry*, 3, 330–333.
- Munro, E. (2004). A simpler way to understand the results of risk assessment instruments. *Children & Youth Services Review*, 26, 873–883. doi:10.1016/j.childyouth.2004.02.026
- Murray, J., & Thomson, M. (2010). Clinical judgement in violence risk assessment. *Europe's Journal of Psychology*, 1, 128-149.
- National Institute for Health and Clinical Excellence. (2009). *Core interventions in the treatment and management of schizophrenia in primary and secondary care*. London, United Kingdom: National Institute for Health and Clinical Excellence.
- Nicholls, T. L., Ogloff, J., Brink, J., & Spidel, A. (2005). Psychopathy in women: A review of its clinical usefulness for assessing risk for aggression and crime. *Behavioral Sciences & the Law*, 23, 779-802.

- Novaco, R. (1994). Anger as a risk factor for violence among the mentally disordered. In J. Monahan & H. J. Steadman (Eds.), *Violence and mental disorder* (pp. 21-59). Chicago, IL: University of Chicago Press.
- Nuffield, J. (1982). *Parole decision making in Canada: Research towards decision guidelines*. Ottawa, ON: Ministry of Supply and Services Canada.
- Nursing and Midwifery Council. (2004). *Standards of proficiency for specialist community public health nurses*. London, United Kingdom: Nursing and Midwifery Council.
- O'Rourke, M., & Hammond, S. (2000). *Risk Management: Towards safe sound and supportive service*. Surrey Hampshire Borders NHS Trust. Retrieved from: <http://www.ramas.co.uk/report1.pdf>
- O'Rourke, M., Hammond, S., Smith, S., & Davies, J. (1998). *Risk Assessment, Management and Audit System: Professional manual*. Surrey: Guilford.
- O'Shea, L. E., & Dickens, G. L. (2014). Short-term Assessment of Risk and Treatability (START): Systematic review and meta-analysis. *Psychological Assessment, 26*, 990-1002. doi: 10.1037/a0036794
- O'Shea, L.E., Picchioni, M.M., & Dickens, G.L. (2015). The predictive validity of the Short-Term Assessment of Risk and Treatability (START) for multiple adverse outcomes in a secure psychiatric inpatient setting. *Assessment*. Advance online publication. doi: 10.1177/1073191115573301
- Ogloff, J., & Davis, M. (2005). Assessing risk for violence in the Australian context. In D. Chappell, & P. Wilson (Eds.), *Issues in Australian crime and criminal justice* (pp. 301-338). Chatswood, Australia: Lexis Nexis Butterworths.
- Olver, M. E., Nicholaichuk, T. P., Gu, D., & Wong, S. C. P. (2013). Sex offender treatment outcome, actuarial risk, and the aging sex offender in Canadian corrections: A long-term follow-up. *Sexual Abuse: A Journal of Research and Treatment, 25*(4), 396-422.

doi: 10.1177/1079063212464399

- Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2014). Thirty years of research on the Level of Service scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment, 26*(1), 156-176.
- Plint, A. C., Moher, D., Morrison, A., Schulz, K. F., Altman, D. G., Hill, C., & Gaboury, I. (2006). Does the CONSORT checklist improve the quality of reports of randomized controlled trials: A systematic review. *Medical Journal of Australia, 185*, 263–267.
- Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing, 41*, 376–382. doi:10.1046/j.1365-2648.2003.02537.x
- Prady, S. L., Richmond, S. J., Morton, V. M., & MacPherson, H. (2008). A systematic evaluation of the impact of STRICTA and CONSORT recommendations on quality of reporting for acupuncture trials. *PLoS ONE, 3*, e1577.  
doi:10.1371/journal.pone.0001577
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk (2nd ed.)*. Washington, DC: American Psychological Association.
- Rettenberger, M., & Eher, R. (2007). Predicting reoffence in sexual offender subtypes: A prospective validation study of the German version of the Sexual Offender Risk Appraisal Guide (SORAG). *Sexual Offender Treatment, 2*, 1-12.
- Rettenberger, M., Hucker, S. J., Boer, D. P., & Eher, R. (2009). The reliability and validity of the *Sexual Violence Risk-20* (SVR-20): An international review. *Sexual Offender Treatment, 4*, 1-14.

- Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting & Clinical Psychology, 63*, 737–748.
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law & Human Behavior, 29*, 615–620. doi:10.1007/s10979-005-6832-7
- Rice, M. E., Harris, G. T., & Lang, C. (2013). Validation of and revision to the VRAG and SORAG: The Violence Risk Appraisal Guide-Revised (VRAG-R). *Psychological Assessment, 25*, 951-965. doi: 10.1037/a0032878
- Risk Management Authority. (2007). *Risk assessment tools evaluation directory (RATED)*. Paisley: Risk Management Authority.
- Roaldset, J., Hartvig, P., & Bjørkly, S. (2011). V-RISK-10: Validation of a screen for risk of violence after discharge from acute psychiatry. *European Psychiatry, 26*, 85–91. doi:10.1016/j.eurpsy.2010.04.002
- Roffey, M., & Kaliski, S. Z. (2012). “To predict or not to predict – that is the question”. An exploration of risk assessment in the context of South African forensic psychiatry. *African Journal of Psychiatry, 15*, 227– 229. doi: 10.4314/ajpsy.v15i4.29
- Rossegger, A., Gerth, J., Urbaniok, F., Laubacher, A., & Endrass, J. (2010). Der Sex Offender Risk Appraisal Guide (SORAG) - Validität und autorisierte deutsche Übersetzung. [The Sex Offender Risk Appraisal Guide (SORAG) - validity and authorized German translation]. *Fortschritte der Neurologie und Psychiatrie, 78*, 658-667. doi: 10.1055/s-0029-1245688
- Rossegger, A., Laubacher, A., Moskvitin, K., Villmar, T., Palermo, G. B., & Endrass, J. (2011). Risk assessment instruments in repeat offending: The usefulness of FOTRES. *International Journal of Offender Therapy and Comparative Criminology, 55*(5), 716-731. doi: 10.1177/0306624X09360662

- Rutherford, M. J., Alterman, A. I., & Cacciola, J. S. (1999). Antisocial personality disorder and psychopathy in cocaine-dependant women. *American Journal of Psychiatry*, *156*, 849-856.
- Rutjes, A. W. S., Reitsma, J. B., Di Nisio, M., Smidt, N., van Rijn, J. C., & Bossuyt, P. M. M. (2006). Evidence of bias and variation in diagnostic accuracy studies. *Canadian Medical Association Journal*, *174*, 469–476. doi:10.1503/cmaj.050090
- Salekin, R. T., Rogers, R., & Sewell, K. W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist-Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice*, *3*(3), 203-215. doi: 10.1111/j.1468-2850.1996.tb00071.x
- Sanderson, S., Tatt, I. D., & Higgins, J. P. T. (2007). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *International Journal of Epidemiology*, *36*, 666–676. doi:10.1093/ije/dym018
- Shih, T. H., & Fan, X. (2009). Comparing response rates in e-mail and paper surveys: A meta-analysis. *Educational Research Review*, *4*, 26–40. doi: 10.1177/1525822208317085
- Shook, J. J., & Sarri, R. C. (2007). Structured decision making in juvenile justice: Judges' and probation officers' perceptions and use. *Children and Youth Services Review*, *29*, 1335–1351. doi: 10.1016/j.chilyouth.2007.05.008
- Simera, I., Moher, D., Hirst, A., Hoey, J., Schulz, K. F., & Altman, D. G. (2010). Transparent and accurate reporting increases reliability, utility, and impact of your research: Reporting guidelines and the EQUATOR Network. *BMC Medicine*, *8*, 24. doi:10.1186/1741-7015-8-24



- Singh, J. P. (2012). The history, development, and testing of forensic risk assessment tools. In E. Grigorenko (Ed.), *Handbook of juvenile forensic psychology and psychiatry*. New York: Springer.
- Singh, J. P. (2013, March). *The International Risk Survey (IRiS) project: Perspectives on the practical application of violence risk assessment tools*. Paper presented at the Annual Conference of the American Psychology-Law Society, Portland, Oregon.
- Singh, J. P. (2013, October). *Violence risk assessment in psychiatry and general medicine*. Speech presented at Yale University, New Haven, CT, USA.
- Singh, J. P. (2013). *The International Risk Survey (IRiS) project: Perspectives on the practical application of violence risk assessment tools*. Paper presented at the Annual Conference of the American Psychology-Law Society, Portland, OR.
- Singh, J. P., & Fazel, S. (2010). Forensic risk assessment: A metareview. *Criminal Justice & Behavior, 37*, 965-988. doi: 10.1177/0093854810374274
- Singh, J. P., Desmarais, S., & Van Dorn, R. A. (2013). Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences & the Law, 31*, 55-73. doi: 10.1002/bsl.2053
- Singh, J. P., Fazel, S., Gueorguieva, R., & Buchanan, A. (2013). Rates of sexual recidivism in high risk sex offenders: A meta-analysis of 10,422 participants. *Sexual Offender Treatment, 7*, 44-57.
- Singh, J. P., Fazel, S., Gueorguieva, R., & Buchanan, A. (2014). Rates of violence in patients classified as high risk by structured risk assessment instruments. *British Journal of Psychiatry, 204*, 180-187. doi: 10.1192/bjp.bp.113.131938
- Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review, 31*, 499-513. doi: 10.1016/j.cpr.2010.11.009

- Singh, J. P., Grann, M., & Fazel, S. (2013). Authorship bias in violence risk assessment? A systematic review and meta-analysis. *PLoS ONE*, *8*, e72484.
- Singh, J. P., Serper, M., Reinharth, J., & Fazel, S. (2011). Structured assessment of violence risk in schizophrenia and other psychiatric disorders: A systematic review of the validity, reliability, and item content of 10 available instruments. *Schizophrenia Bulletin*, *37*(5), 899-912. doi: 10.1093/schbul/sbr093
- Singh, J. P., Yang, S., Mulvey, E., & the RAGEE Group. (in press). Reporting standards for risk assessment predictive validity studies: The Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement. *Law & Human Behavior*.
- Sjöstedt, G., & Grann, M. (2002). Risk assessment: What is being predicted by actuarial prediction instruments? *International Journal of Forensic Mental Health*, *1*, 179–183. doi:10.1080/14999013.2002.10471172
- Sjöstedt, G., & Långström, N. (2003). 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. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science*, *20*, 38–42. doi:10.1177/0963721410397271
- Skelton, A., & Vess, J. (2008). Risk of sexual recidivism as a function of age and actuarial risk. *Journal of Sexual Aggression*, *14*(3), 199-209. doi: 10.1080/13552600802267098
- Slovenko, R. (1975). Psychotherapy and confidentiality. *Cleveland State Law Review*, *24*, 375-391.
- Smallbone, S., & Rallings, M. (2013). Short-term predictive validity of the Static-99 and Static-99-R for indigenous and nonindigenous Australian sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *25*(3), 302-316. doi: 10.1177/1079063212472937

- Smidt, N., Rutjes, A. W., van der Windt, D. A., Ostelo, R. W., Bossuyt, P. M., Reitsma, J. B., . . . de Vet, H. C. (2006). The quality of diagnostic accuracy studies since the STARD statement: Has it improved? *Neurology*, *67*, 792–797.  
doi:10.1212/01.wnl.0000238386.41398.30
- Smith, B. A., Lee, H. J., Lee, J. H., Choi, M., Jones, D. E., Bausell, R. B., & Broome, M. E. (2008). Quality of reporting randomized controlled trials (RCTs) in the nursing literature: Application of the Consolidated Standards of Reporting Trials (CONSORT). *Nursing Outlook*, *56*, 31–37. doi:10.1016/j.outlook.2007.09.002
- Smits, N. (2010). A note on Youden's J and its cost ratio. *BMC Medical Research Methodology*, *10*, 89–93. doi:10.1186/1471-2288-10-89
- Sreenivasan, S., Garrick, T., Norris, R., Cusworth-Walker, S., Weinberger, L. E., Essres, G., et al. (2007). Predicting the likelihood of future sexual recidivism: Pilot study findings from a California sex offender risk project and cross-validation of the Static-99. *Journal of the American Academy of Psychiatry & the Law*, *35*, 454–468.
- StataCorp (2007). *STATA statistical software: Release 10.1*. College Station, TX: StataCorp LP.
- Steadman, H. J., & Cocozza, J. J. (1974). *Careers of the criminally insane: Excessive social control of deviance*. Lexington, MA: DC Health.
- Steeg, S., Kapur, N., Webb, R., Applegate, E., Stewart, S. L. K., Hawton, K., . . . Cooper, J. (2012). *Psychological Medicine*, *42*, 2383-2394. doi: 10.1017/S0033291712000347
- Storey, J. E., Watt, K. A., Jackson, K. J., & Hart, S. D. (2012). Utilization and implications of the Static-99 in practice. *Sexual Abuse: A Journal of Research and Treatment*, *24*(3), 289-302. doi: 10.1177/1079063211423943

- Studer, L. H., Aylwin, A. S., Stribney, C., & Reddon, J. R. (2011). *Uses, misuses, and abuses of risk assessment with sexual offenders*. In D. P. Boer, R., Eher, L. A. Craig, M. H. Miner, & F. Pfäfflin (Eds.). *International Perspectives on the Assessment and Treatment of Sexual Offenders: Theory, Practice and Research* (pp. 193-212). Chichester, United Kingdom: Wiley.
- Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*, 1285–1293. doi:10.1126/science.3287615
- Swinburne Romine, R. E., Miner, M. H., Poulin, D., Dwyer, S. M., & Berg, D. (2012). Predicting reoffense for community-based sexual offenders: An analysis of 30 years of data. *Sexual Abuse: A Journal of Research and Treatment*, *24*(5), 501-514. doi: 10.1177/1079063212446514
- Tape, T. (2006). *The area under the ROC curve*. In *Interpreting diagnostic tests*. Retrieved from <http://gim.unmc.edu/dxtests/ROC3.htm>
- Taxman, F. S., Cropsey, K. L., Young, D. W., & Wexler, H. (2007). Screening, assessment, and referral practices in adult correctional settings: A national perspective. *Criminal Justice & Behavior*, *34*, 1216– 1234. doi: 10.1177/0093854807304431
- Telles, L. E., Day, V. P., Folino, J. O., & Taborda, J. G. (2009). Reliability of the Brazilian version of HCR-20 Assessing Risk for Violence. *Revista Brasileira de Psiquiatria*, *31*, 253–256.
- Thomson, L., Davidson, M., Brett, C., Steele, J., & Darjee, R. (2008). Risk assessment in forensic patients with schizophrenia: The predictive validity of actuarial scales and symptom severity for offending and violence over 8–10 years. *International Journal of Forensic Mental Health*, *7*, 173–189. doi:10.1080/ 14999013.2008.9914413
- Thornberry, T., & Jacoby, J. E. (1979). *The criminally insane: A community follow-up of mentally ill offenders*. Chicago, IL: University of Chicago Press.

- Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). *Distinguishing and combining risks for sexual and violent recidivism*. In R. Prentky, E. Janus, M. Seto, and A.W. Burgess (Eds), *Understanding and managing sexually coercive behavior. Annals of the New York Academy of Sciences*, 989, 225–235.
- Tolman, A., & Mullendore, K. (2003). Risk evaluations for the courts: Is service quality a function of specialization? *Professional Psychology: Research & Practice*, 34, 225–232. doi:10.1037/0735-7028.34.3.225
- Troquete, N. A., van den Brink, R. H. S., Beintema, H., Mulder, T., van Os, T. W., & Schoevers, R. A. (2013) Risk assessment and shared care planning in out-patient forensic psychiatry: Cluster randomised controlled trial. *British Journal of Psychiatry*, 202, 365–371. doi:10.1192/bjp. bp.112.113043
- Tully, R.J. & Browne, K.D. (2015). Appraising the Risk Matrix 2000 static sex offender risk assessment tool. *International Journal of Offender Therapy and Comparative Criminology*, 59(2), 211-214.
- Tyrer, P., Duggan, C., Cooper, S., Crawford, M., Seivewright, H., Rutter, D., et al. (2010). The successes and failures of the DSPD experiment: The assessment and management of severe personality disorder. *Medicine, Science and the Law*, 50, 95–99. doi:10.1258/msl.2010.010001
- Underwood, D., Kim, H., & Matier, M. (2000, May). *To mail or to Web: Comparisons of survey response rates and respondent characteristics*. Paper presented at the 40th Annual Forum of the Association for Institutional Research, Cincinnati, OH.
- Urbaniok, F. (2007). *FOTRES: Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System* [Forensic Operationalized Therapy/Risk Evaluation-System]. Bern: Zytglogge.

- Urbaniok, F. (2009). FOTRES: Forensic Operationalized Therapy/Risk Evaluation-System (Version 2.0) [Computer software]. Retrieved from <http://www.fotres.ch>
- Van Marle, H. J. C. (2002). The Dutch Entrustment Act (TBS): Its principles and innovations. *International Journal of Forensic Mental Health, 1*(1), 83-92. doi: 10.1080/14999013.2002.10471163
- Viljoen, J. L., McLachlan, K., & Vincent, G. M. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment, 17*, 377–395. doi: 10.1177/1073191109359587
- Viljoen, J., Scalora, M., Cuadra, L., Bader, S., Chavez, V., Ullman, D., & Lawrence, L. (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 & Behavior, 35*(1), 5-23.
- Vitale, J. E., & Neumann, J. P. (2001). Using the Psychopathy Checklist-Revised with female samples: Reliability, validity, and implications for clinical utility. *Clinical Psychology : Science and Practice, 8*, 117-132.
- Wakeling, H. C., Howard, P., & Barnett, G. (2011). Comparing the validity of the RM2000 scales and OGRS3 for predicting recidivism by internet sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 23*(1), 146-168. doi: 10.1177/1079063210375974
- Walters, G. D. (2003). Predicting institutional adjustment and recidivism with the Psychopathy Checklist factor scores: A meta-analysis. *Law & Human Behavior, 27*(5), 541. doi: 10.1023/A:1025490207678
- Warren, J. I., Burnette, M., South, S. C., Chauhan, P., Bale, R., Friend, R., & Van Patten, I. (2003). Psychopathy in women: Structural modeling and comorbidity. *International Journal of Law and Psychiatry, 37*(2), 1-21.

- Watt, K. A., & Jackson, K. (2008, July). *Interrater and structural reliabilities of the Risk for Sexual Violence Protocol (RSVP)*. Paper presented at the annual meeting of the International Association of Forensic Mental Health Services, Vienna, Austria.
- Waypoint Centre for Mental Health Care. (2014). *Research department bibliography on assessment and communication of violence risk*. Retrieved from <http://static.squarespace.com/static/520a76a0e4b03ad27abae1e3/t/53f52df0e4b07b3557c47122/1408577008298/Risk.pdf>
- Webster, C. D., & Belfrage, H. (2013). *HCR-20 (Version 3): Assessing Risk for Violence*. Burnaby, BC, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Webster, C. D., Douglas, K., S., Eaves, D. & Hart, S. D. (1997). *HCR-20: Assessing the Risk for Violence (Version 2)*. Burnaby, BC: Simon Fraser University, Mental Health Law and Policy Institute.
- Webster, C. D., Eaves, D., Douglas, K. S., & Wintrup, A. (1995). *The HCR-20 scheme: The assessment of dangerousness and risk*. Vancouver, BC: Mental Health Law and Policy Institute, and Forensic Psychiatric Services Commission of British Columbia.
- Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. (2009). *Manual for the Short-Term Assessment of Risk and Treatability (START). Version 1.1*. Hamilton, ON: Forensic Psychiatric Services Commission.
- 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 violence risk structured professional judgment scheme. *Behavioral Sciences & the Law*, 24, 747–766.
- Wechsler, D. (1997). *The Wechsler Adult Intelligence Scale (3rd ed.)*. San Antonio, TX: The Psychological Corporation.

- Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children (4th ed.)*. San Antonio, TX: The Psychological Corporation.
- Welsh, J., Schmidt, F., McKinnon, L., Cattha, 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. [Findings of a Nationwide Pilot Study on the HKT-30]*. The Hague, Netherlands: Ministerie van Justitie.
- Whittington, R., Bjørngaard, J., Brown, A., Nathan, R., Noblett, S., & Quinn, B. (2014). Dynamic relationship between multiple START assessments and violent incidents over time: A prospective cohort study. *BMC Psychiatry*. Advance online publication. doi: 10.1186/s12888-014-0323-7
- Wilcox, D., Beech, A., Markall, H. F., & Blacker, J. (2009). Actuarial risk assessment and recidivism in a sample of UK intellectually disabled sexual offenders. *Journal of Sexual Aggression, 15*(1), 97-106. doi: 10.1080/13552600802578577
- Wilson, R. J., Abracen, J., Looman, J., Picheca, J. E., & Ferguson, M. (2011). Pedophilia: An evaluation of diagnostic and risk prediction methods. *Sexual Abuse: A Journal of Research and Treatment, 23*(2), 260-274. doi: 10.1177/1079063210384277
- Wong S, & Gordon A. (2009) *Manual for the Violence Risk Scale*. Saskatoon, SK: University of Saskatchewan.
- Woods, P., & Ashley, C. (2007). Violence and aggression: A literature review. *Journal of Psychiatric & Mental Health Nursing, 14*, 652–660. doi:10.1111/j.1365-2850.2007.01149.x



- Wootton, L., Buchanan, A., Leese, M., Tyrer, P., Burns, T., Creed, F., et al. (2008). Violence in psychosis: Estimating the predictive validity of readily accessible clinical information in a community sample. *Schizophrenia Research, 101*, 176–184.  
doi:10.1016/j.schres.2007.12.490
- World Health Organization. (2002). *World report on violence and health*. Geneva: Author.
- Worling, J. R., Bookalam, D., & Litteljohn, A. (2012). Prospective validity of the Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR). *Sexual Abuse: A Journal of Research and Treatment, 24*(3), 203-223. doi: 10.1177/1079063211407080
- Yang, M., Wong, S. C., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*, 740-767. doi:10.1037/a0020473
- Zanatta, R. (2005). *Risk of violence and sexual recidivism: A comparison of Dangerous Offenders and repetitive sexual offenders*. Unpublished doctoral dissertation. Department of Psychology, Simon Fraser University, Burnaby, BC, Canada.
- Zhang, X. L., Chen, X. C., Cai, W. X., Hu, J. M. (2012). Reliability of the Violence Risk Scale Chinese version. *Fai Yi Xue Za Zhi, 28*, 32–35.
- Zucker, K. J. (2013). The impact factor: Just the facts. *Archives of Sexual Behavior*.  
doi:10.1007/s10508-013-0105-1

## **RAGEE SUPPLEMENT**

---

**R1 (Abstract).** Include a structured abstract describing the study

**Objectives:** An actuarial model was developed in the MacArthur Violence Risk Assessment Study to predict violence in the community among patients who have recently been discharged from psychiatric facilities. This model, called the multiple iterative classification tree (ICT) model, showed considerable accuracy in predicting violence in the construction sample. The purpose of the study reported here was to determine the validity of the multiple ICT model in distinguishing between patients with high and low risk of violence in the community when applied to a new sample of individuals.

**Methods:** Software incorporating the multiple ICT model was administered with independent samples of acutely hospitalized civil patients. Patients who were classified as having a high or a low risk of violence were followed in the community for 20 weeks after discharge. Violence included any battery with physical injury, use of a weapon, threats made with a weapon in hand, and sexual assault.

**Results:** Expected rates of violence in the low- and high-risk groups were 1 percent and 64 percent, respectively. Observed rates of violence in the low- and high-risk groups were 9 percent and 35 percent, respectively, when a strict definition of violence was used, and 9 percent and 49 percent, respectively, when a slightly more inclusive definition of violence was used. These findings may reflect the “shrinkage” expected in moving from construction to validation samples.

**Conclusions:** The multiple ICT model may be helpful to clinicians who are faced with making decisions about discharge planning for acutely hospitalized civil patients.

Monahan, J, Steadman, H., Robbins, P., Appelbaum, P., Banks, S., Grisso, T., Heilbrun, K., Mulvey, E., Roth, L., & Silver, E. (2005). An actuarial model of violence risk assessment for persons with mental disorders. *Psychiatric Services*, 56, 810-815.

Comment: A brief description including the study aims, sample, methods, principal results, and implications should be reported, in keeping with the specific journal’s requirements.

**R2 (Abstract).** Identify the article as a risk assessment study in which predictive validity is measured

**Background:** Current violence risk assessment instruments are time-consuming and mainly developed for forensic psychiatry. A paucity of violence screens for acute psychiatry instigated the development and validation of the V-RISK-10. **The aim of this prospective naturalistic study was to test the predictive validity of the V-RISK-10 as a screen of violence risk after discharge from two acute psychiatric wards.**

**Methods:** Patients were screened with V-RISK-10 before discharge, and incidents of violence were recorded 3, 6, 9 and 12 months after discharge. A total of 381 of the 1017 patients that were screened completed the follow up.

**Results:** The ROC-AUC values for any violent behaviour were 0.80 and 0.75 ( $p < 0.001$ ) for the 3 and 12 months follow-up periods, respectively, and significant for both genders. The most accurate risk estimates were obtained for severe violence. For persons without a known history of violence prior to the screening, AUCs were 0.74 ( $p = 0.004$ ) and 0.68 ( $p = 0.002$ ).

**Conclusions:** Results indicate that the V-RISK-10 is a valid and clinically useful screen for violence risk after discharge from acute psychiatry, and even significant for patients without a known previous history of violence.

Roaldset, J. O., Hartvig, P., & Bjørkly, S. (2011). V-RISK-10: Validation of a screen for risk of violence after discharge from acute psychiatry. *European Psychiatry*, 26, 85-91.

Comment: Predictive validity is defined as the extent to which an instrument-based estimate (e.g., total risk score, actuarial risk bin, or final risk judgment) predicts an outcome measure (e.g., arrest, conviction, or incarceration for a violent offense).

**R3 (Abstract).** Identify the risk assessment instrument(s) whose predictive validity is measured

**Objective:** This study tested the interrater reliability and criterion-related validity of structured violence risk judgments made by using one application of the structured professional judgment model of violence risk assessment, **the HCR-20 violence risk assessment scheme**, which assesses 20 key risk factors in three domains: historical, clinical, and risk management.

**Methods:** The HCR-20 was completed for a sample of 100 forensic psychiatric patients who had been found not guilty by reason of a mental disorder and were subsequently released to the community. Violence in the community was determined from multiple file-based sources.

**Results:** Interrater reliability of structured final risk judgments of low, moderate, or high violence risk made on the basis of the structured professional judgment model was acceptable (weighted kappa=.61). Structured final risk judgments were significantly predictive of postrelease community violence, yielding moderate to large effect sizes. Event history analyses showed that final risk judgments made with the structured professional judgment model added incremental validity to the HCR-20 used in an actuarial (numerical) sense.

**Conclusions:** The findings support the structured professional judgment model of risk assessment as well as the HCR-20 specifically and suggest that clinical judgment, if made within a structured context, can contribute in meaningful ways to the assessment of violence risk.

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.

Comment: Report the acronym(s) and/or full name(s) of the violence risk assessment instrument(s) whose predictive validity was examined in the study.

**R4 (Abstract).** State the nature of the principal outcome (e.g., violence, sexual violence, criminal offending, institutional misconduct)

**Purpose:** To investigate the validity of risk factors and established risk measures in predicting community violence in an acute mental health sample up to 20 weeks post-discharge.

**Method:** Prospective cohort follow-up study conducted between January 2006 and August 2007. Baseline assessments were conducted while participants were inpatients. The measures were rated following interview with the participants, record review and speaking to someone who knows the person well (e.g. friend, relative, carer). Baseline measures were then compared with frequency and severity of **violence in the community** post-discharge at 20 weeks.

**Results:** In the 20-week period post-discharge, 29 (25.4%) of the 114 participants were violent. All the risk measures and measures of impulsiveness and anger were predictive of violence where  $p < 0.05$ . The HCR-20 total, psychopathy and clinical factors were strongly correlated with the frequency of violence where  $p < 0.05$ .

**Conclusions:** The risk factors and risk measures that have been found to be predictive in forensic samples are also predictive in acute mental health samples, although the effects are not as large. Future research needs to be conducted with a larger sample to include investigation of differences in risk factors based on gender and social support. Services and clinicians need to consider how to integrate findings into useful frameworks to support decisions and contribute to managing risk. This should assist in identifying interventions aimed at preventing community violence.

Doyle, M., Carter, S., Shaw, J., & Dolan, M. (2011). Predicting community violence from patients discharged from acute mental health units in England. *Social Psychiatry & Psychiatric Epidemiology*, 47, 627-637.

Comment: The principal outcome is defined as that which was used as the primary dependent variable in predictive validity analyses.

**R5 (Introduction).** Provide the rationale and a summary of the scientific/theoretical background for the study

START has experienced quick uptake into clinical practice: We are aware of implementations in at least 10 countries, and the manual has been translated into four languages, with an additional four translations underway. **However, only a handful of studies have examined the reliability and validity of START assessments (Braithwaite, Charette, Crocker, & Reyes, 2010; Chu et al., 2011; Gray et al., in press; Nicholls, Brink, Desmarais, Webster, & Martin, 2006; Nonstad et al., 2010). The original validation study was published by Nicholls et al. (2006). Evaluating START assessments completed by nurses, social workers, and psychiatrists regarding 137 male forensic psychiatric patients, Nicholls et al. found excellent interrater agreement overall (intraclass correlation coefficient,  $ICC_2 = .87, p < .001$ ) and within professional disciplines nursing = .88; social work = .92; and psychiatry = .80; all  $ps < .001$ ). The authors also reported significantly higher START total scores for patients who engaged in aggression over the 12-month follow-up: any aggression to others ( $M = 75.66$  vs. 65.86), verbal aggression ( $M = 75.86$  vs. 66.82), aggression against objects ( $M = 77.90$  vs. 68.00), physical aggression against others ( $M = 76.32$  vs. 68.25), violence against others ( $M = 81.82$  vs. 69.12), and sexual aggression ( $M = 80.63$  vs. 70.24; all  $ps < .05$ ). Receiver operating characteristic (ROC) analyses of a subsample of 50 patients who remained hospitalized throughout follow up revealed good validity in predicting verbal aggression (area under the curve [AUC] = .72,  $SE = 0.07, p < .001$ ), physical aggression against objects (AUC = .67,  $SE = 0.08, p < .05$ ), physical aggression against others (AUC = .70,  $SE = 0.08, p < .01$ ), and sexually inappropriate behavior (AUC = .92,  $SE = 0.10, p < .05$ ).**

Results of this research are promising, but further evaluation is necessary for several reasons. The predictive validity of the final risk estimates, one of the identifying features of the structured professional judgment approach (Singh, Grann, & Fazel, 2011; Skeem & Monahan, 2011), has only been examined in two studies (Braithwaite et al., 2010; Gray et al., in press), and the validation samples have been quite small ( $ns = 34-50$ ). Furthermore, there have been several significant changes to the instrument since these evaluations. Now, each of the 20 START items is scored in terms of *both* vulnerability and strength, and final risk estimates of low, moderate, or high are made across seven outcome domains (Webster et al., 2004). In 2009, the authors published a text revision of the START manual (Version 1.1; Webster et al., 2009), which included content updates to three items (Mental State, Emotional State, and Treatability), explicit operationalization of START components left undefined in the 2004 consultation edition of the manual (e.g., each of the risk outcome domains, strengths and vulnerabilities, and key and critical items), and specification of coding time frames (i.e., item ratings based on functioning over the past 2 to 3 months or since the previous START assessment).

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

Comment: The rationale for the study is defined as the reason why the present investigation is necessary given existing evidence. Describing the scientific/theoretical background for the study serves to situate the investigation in light of previous research, contemporary legal statutes, organizational and/or government reports, and relevant clinical guidelines.

**R6 (Introduction).** State the research question(s) and/or study aim(s)

The present study is part of a large-scale research project concerning the clinical, criminological, and legal aspects of sexual murderers using a comprehensive and comparatively large sample of offenders convicted of sexually motivated murder in Germany (e.g., Berner et al, 2008; Briken, Habermann, Kafka, Berner, & Hill, 2006; Briken, Nika, & Berner, 1999; Hill, Habermann, Berner, & Briken, 2006; Hill, Habermann et al., 2008; Hill, Ujeyl et al., 2008; Ujeyl et al., 2008). **The main aim of the present study was to examine the predictive accuracy of four well established risk assessment instruments: the Static-99 (Hanson & Thornton, 2000), the Historical-Clinical-Risk Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997), the Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster., 1997), and the Psychopathy Checklist-Revised (PCL-R; Hare, 2003).** In order to achieve comparability with the existing status of risk assessment research we used commonly used effect sizes for the investigation of the predictive accuracy of the instruments. **Furthermore, we examined different recidivism criteria in order to prove differential effects, and we investigated the predictive accuracy of the different subscales and the individual items of the instruments as well.**

Hill, A., Rettenberger, M., Habermann, N., Berner, W., Eher, R., & Briken, P. (2012). The utility of risk assessment instruments for the prediction of recidivism in sexual homicide perpetrators. *Journal of Interpersonal Violence, 27*, 1-26.

Comment: Identify the main points of inquiry that the present research sought to answer. Whereas the rationale and scientific/theoretical background serve to provide larger context within the field, the research question(s) and specific aims(s) focus more precisely on the particular issues addressed in the present study.



**R7 (Methods).** Report the sample size

Participants were **99** male adolescents who had been released from custody. They had been at liberty for a minimum of 12 months and were traced on the Home Office Police National Computer (HOPNC). The 99 participants constituted 80.5% of a baseline sample of 123 male adolescents who had been assessed on measures of personality and risk while in custody.

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.

Comment: Identify the number of participants in the study.

**R8 (Methods).** Report the sex/gender composition of the sample

Between April 15, 2005 and December 31, 2007, the clinical staff completed 258 START assessments of 61 patients. Of these, 47 could be included in the validation study because they had been in the hospital for three months after their first START assessment. **These were 39 men (83%) and eight women (17%).**

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

Comment: Provide the number and/or percentage of the participants who were men/male and/or women/female.

**R9 (Methods).** Report the average age at assessment (with dispersion parameter)

**Average age at discharge from inpatient forensic psychiatric care, release from prison or onset of probation was 34.07 (SD = 11.00, range 16-58) years.** ... The study had a retrospective follow-up design. Subjects were followed from release until first event of a sexual or a violent non-sex reoffence resulting in a new criminal conviction, or the end of follow-up (June 1<sup>st</sup>, 1999).

Sjöstedt, G., & Långström, N. (2002). Assessment of risk for criminal recidivism among rapists: A comparison of four different measures. *Psychology, Crime & Law*, 8, 25-40.

Comment: The mean age should be reported in years. A dispersion parameter should be reported to describe spread, using a standard deviation or confidence interval. In cases where the distribution is skewed, a median and interquartile range may be substituted.

**R10 (Methods).** Report the index offense composition of the sample

**The breakdown of the sample by current offence at the start of sentence was as follows: 61 violent (28.2%), 28 burglary (13.0%), 21 theft (9.7%), 15 sexual (6.9%), 7 fraud and forgery (3.2%), 3 criminal damage (1.4%), and 81 (36.1%) offenders were convicted for a diverse range of offences, including motoring, drugs, and financial crimes.**

Dowdy, E. R., Lacy, M. G., Unnithan, N. P. (2002). Correctional prediction and the Level of Supervision Inventory. *Journal of Criminal Justice*, 30, 29-39.

Comment: When the study sample involves a correctional and/or forensic population, provide the number and/or percentage of the participants who engaged in each type of act that resulted in contact with the criminal justice or forensic mental health setting where risk assessments took place. This item may not be relevant for non-forensic samples.

**R11 (Methods).** Report the characteristics of groups that underwent subgroup analysis

<b>Variable</b>	<b>MIC (n = 101)</b>	<b>MIOF (n = 100)</b>	<b>Jail (n = 107)</b>	<b>Total (n = 308)</b>
<b>Race and ethnicity</b>				
<b>African American</b>	<b>18 (17.8)</b>	<b>14 (14.0)</b>	<b>28 (26.2)</b>	60 (19.5)
<b>Hispanic</b>	<b>54 (53.5)</b>	<b>46 (46.0)</b>	<b>48 (44.9)</b>	147 (47.9)
<b>White</b>	<b>29 (28.7)</b>	<b>40 (40.0)</b>	<b>31 (29.0)</b>	100 (32.6)
<b>Gender</b>				
<b>Male</b>	<b>44 (43.6)</b>	<b>71 (71.0)</b>	<b>61 (57.0)</b>	175 (57.0)
<b>Female</b>	<b>57 (56.4)</b>	<b>29 (29.0)</b>	<b>46 (43.0)</b>	132 (43.0)
<b>Age</b>				
<b>18-39</b>	<b>59 (58.4)</b>	<b>83 (83.0)</b>	<b>68 (63.6)</b>	210 (68.4)
<b>40-61</b>	<b>42 (41.6)</b>	<b>17 (17.0)</b>	<b>39 (46.4)</b>	97 (31.6)
<b>Marital status</b>				
<b>Single or divorced</b>	<b>86 (85.1)</b>	<b>83 (83.0)</b>	<b>97 (90.7)</b>	265 (86.3)
<b>Married</b>	<b>15 (14.9)</b>	<b>17 (17.0)</b>	<b>10 (9.3)</b>	42 (13.7)

Note: Values are *n* (%). MIC = Mentally Impaired Caseload; MIOF = Mentally Impaired Offender Facility; SSI = Supplemental Security Income; Unavail = Data not

Castillo, D. E., & Alarid, L. F. (2011). Factors associated with recidivism among offenders with mental illness. *International Journal of Offender Therapy & Comparative Criminology*, 55, 98-117.

Comment: Provide the same (or a relevant subset of) descriptive characteristics for groups that underwent subgroup analysis as were provided for the overall sample. This may be done in the text or in a table as in the example above.

**R12 (Methods).** Report the acronym(s) and full name(s) of the instrument(s) under investigation with appropriate reference to source document

**The Psychopathy Checklist-Revised (PCL-R; Hare, 1991)** contains 20 items, each scored on a 3-point scale from 0 to 2, giving a total score ranging from 0 to 40. It was developed as a measure of the extent to which an individual matches Cleckley's (1941) description of the prototypical psychopath, and has been found to be a good predictor of violent recidivism (Dolan & Doyle, 2000).

Thomson, L., Davidson, M., Brett, C., Steele, J., & Darjee, R. (2008). Risk assessment in forensic patients with schizophrenia: The predictive validity of actuarial scales and symptom severity for offending and violence over 8-10 years. *International Journal of Forensic Mental Health*, 7, 173-189.

Comment: A source document is a report that initially presents an instrument's items, discusses its appropriate use, and provides information on psychometric validation. Appropriate source documents include articles, books, government reports, Masters theses, doctoral dissertations, and conference presentations, with preference given to published manuals and peer-reviewed instrument development studies.

**R13 (Methods).** Report the number of items on the instrument(s) under investigation

The VRAG (Quinsey et al., 1998) **comprises 12 items**, including such items as the Psychopathy Checklist Revised (PCL- R; Hare, 2004) score (which in turn has 20 items), elementary school adjustment, offender's age at time of offence, etc. If we could not score a particular item then that item was rated as a '0'. We note that the updated manual (Quinsey et al., 2006) pro-rating is now recommended. We did not score the VRAG if more than four items could not be scored.

Snowden, R. J., Gray, N. S., Taylor, J., & MacCulloch, M. J. (2007). Actuarial prediction of violent recidivism in mentally disordered offenders. *Psychological Medicine*, 37, 1539-1549.

Comment: This value should represent the total number of originally published items on each instrument whose predictive validity was measured in the study. Modifications to the instrument(s), such as systematically omitted items, should be reported separately.

**R14 (Methods).** Report the approach by which the assessment information from the instrument(s) under investigation is organized into an overall evaluation of risk

*J-SORRAT-II.* The J-SORRAT-II is a 12-item **actuarial tool** designed for assessing risk of violence among male juvenile offenders who were 12 to 18 years old at the time of their index sexual offense (Epperson et al., 2005). A number of items on the J-SORRAT-II focus on the youths' sexual and nonsexual offense history (e.g., number of adjudications as a sex offender, number of victims in sex offenses). Other variables examine youths' treatment history (i.e., completion of sex offender treatment), school history (e.g., special education), and past victimization experiences (e.g., number of physical abuse victimization events).

Viljoen, J. L., Scalora, M., Cuadra, L., Bader, S., Chavez, V., Ullman, D., & Lawrence, L. (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 & Behavior*, 35, 5-23.

Comment: Common approaches in forensic risk assessment combine information gathered at the item level using one of two strategies to arrive at a risk estimate. The first of these approaches is actuarial assessment, which involves the estimation of the likelihood of future antisocial behavior by assigning numerical values to factors associated with offending. These numbers are then combined using a statistical algorithm to translate an individual's total score into a group-based probabilistic estimate of future antisocial behavior. The second approach is referred to as structured professional judgment (SPJ), which involves *aide-mémoires* that guide assessors to estimate risk after reviewing empirically- and theoretically-based risk and/or protective factors.



**R15 (Methods).** Report the population for which the instrument(s) under investigation was intended to be used

The START is a structured professional judgment guide for the assessment of seven often interrelated risks associated with mental, substance use, and personality disorders in adults: violence to others, self-harm, suicide, unauthorized leave, substance abuse, self-neglect, and being victimized. The instrument consists of 20 dynamic factors that are assessed for both Strength and Vulnerability on a 3-point ordinal scale from 0 (minimally present) to 2 (maximally present). Strength and Vulnerability ratings should be scored independent of one another, and a patient may be scored high (or low) on both Strength and Vulnerability for any particular item. For example, a patient may receive a high Vulnerability rating for relationships (Item 2) if he or she is involved in an abusive intimate relationship but also may receive a high Strength rating if he or she has a warm, loving, and reciprocal relationship with his or her parents, other family members, or peers. Based on item ratings, identification of key and critical items (i.e., items that are particularly relevant, either recently or historically, to individual risk), and consideration of historical factors, assessors estimate risk as low, moderate, or high for each of the seven outcome domains. Strength and Vulnerability total scores can be calculated for research purposes by summing the item ratings (possible range = 0–40). **START is intended for use with both inpatient and outpatient populations in civil psychiatric, forensic psychiatric, and correctional settings.**

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

Comment: Relevant description of the population may include: setting (institutional/inpatient vs. community/outpatient), forensic status (forensic vs. non-forensic), age (adult vs. juvenile), sex (men vs. women), and whether the instrument was designed to be used in mental health and/or correctional contexts.

**R16 (Methods).** Report the outcome(s) that the instrument(s) under investigation was intended to assess

The Violence Risk Appraisal Guide (VRAG) is an actuarial violence risk assessment developed on 618 violent offenders evaluated in a maximum security forensic psychiatric facility. Most in this development sample were convicted before or after the evaluation while a minority was found not guilty by reason of insanity; about a quarter met the diagnostic criteria for a psychotic disorder. In development, the VRAG's items were selected for their ability to provide independent and incremental information about the likelihood that subjects later met **the operational definition of violent recidivism – a criminal charge for a violent offense or reinstitutionalization for violent conduct that would otherwise have resulted in a criminal charge.**

Harris, G. T., & Rice, M. E. (2007). Adjusting actuarial violence risk assessments based on aging or the passage of time. *Criminal Justice & Behavior*, 34, 297-313.

Comment: Provide both the type of outcome (e.g., violent, sexual, general offending) as well as the source of outcome detection (e.g., criminal charges, arrest, conviction, incarceration, self-report, collateral interviews). As risk assessment instruments other than those adopting the actuarial approach may not have been intended to assess the likelihood of an outcome detected via a particular source, the latter criterion may not be relevant for all risk assessment instruments.

**R17 (Methods).** Report the length of follow-up for which manual-recommended probability estimates of risk were derived for the instrument(s) under investigation

The SORAG is an actuarial risk assessment tool for sexual offenders that was developed by Canadian forensic researchers. This instrument is a modification of the Violence Risk Appraisal Guide (VRAG; Quinsey et al., 2006), which was developed to predict violent and sexual recidivism among adult male offenders; 10 of the 14 items of the SORAG are the same as in the VRAG. The SORAG is conceptualized for sexual offenders to assess violent recidivism risk, which includes sexual offences involving physical contact with the victim. The instrument consists of 14 weighted items: lived with biological parents up to age 16, elementary school maladjustment, history of alcohol problems, marital status, criminal history for nonviolent offences, criminal history for violent offences, previous convictions for sexual offences, sexual offences against girls under age 14 only, failure on prior conditional releases, age at index offence, *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed.) (American Psychiatric Association, 1980) criteria for any personality disorder, *DSM-III* criteria for schizophrenia, phallometric test results indicating pedophilia or sexual sadism, and PCL-R score. Based on the total score the evaluator can allocate the offender to one of nine risk categories. **By means of these risk categories, it is possible to infer to empirically calculated probabilities of violent (including sexual) recidivism after 7 and 10 years at risk, respectively.**

Rettenberger, M., Matthes, A., Boer, D. P., & Eher, R. (2010). Prospective actuarial risk assessment: A comparison of five risk assessment instruments in different sexual offender subtypes. *International Journal of Offender Therapy & Comparative Criminology*, 54, 169-186.

Comment: This information is routinely available in instrument manuals and development studies. This item may not be relevant for non-actuarial risk assessment instruments, as the generation of probabilistic risk estimates is unique to the actuarial approach.

**R18 (Methods).** Report the cut-off score(s) and/or risk categories that the instrument(s) under investigation was designed to use to classify risk level

The Static-99 is composed of 10 historical risk factors (see Table 2) that have to be coded from file information. **The factors add up to a maximum total score of 12 that is subsequently translated into four risk categories: low (0,1), medium low (2-3), medium high (4-5) and high (6 or more; Hanson & Thornton, 1999).**

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 & Human Behavior*, 28, 235-251.

Comment: Cut-off scores and risk categories established by the developers of the instrument(s) should be reported. If the study used different cut-off scores and/or risk categories than those identified by the instrument developers, this should also be stated.

**R19 (Methods).** Report whether risk assessments were conducted in the context of research or practice

**One hundred thirty-two youth in custody were invited to participate in the study. Of these, parent/legal guardians refused consent for 28 youth (21%), five youth refused consent (4%), and one youth withdrew partway through the study (<1%). Furthermore, 19 of these youth (14%) did not receive SAVRY ratings due to insufficient collateral information to code the measure or because the youth did not complete a research interview. We invited 102 youth from the mental health assessment center to participate. Of these, 19 youth refused consent (19%) and two youth withdrew partway through the study (2%). Sixteen of these youth (16%) did not receive SAVRY ratings for the same reasons listed above. The gender and age composition of youth who did not participate in the study was not significantly different from youth who consented to participate (for gender,  $\chi^2 = 0.31, P > .05$ ; for age,  $F(1,226) = .78, P > .05$ ).**

Penney, S. R., Lee, Z., & Moretti, M. M. (2010). Gender differences in risk factors for violence: An examination of the predictive validity of the Structured Assessment of Violence Risk in Youth. *Aggressive Behavior, 36*, 390-404.

Comment: Studies conducted in a research context are those in which the risk assessment instruments under investigation *were not* administered as part of routine practice. Studies conducted in a practice context are those in which the risk assessment instruments under investigation *were* administered as part of routine practice. As the example above suggests, this criterion can be met by stating that risk assessments were conducted as part of a voluntary research process.

**R20 (Methods).** Identify when risk assessments occurred (e.g., pre-admission, admission, release, post-release)

The HCR–20 consists of 20 items: 10 items related to historical factors (e.g. employment problems, history of mental illness), 5 items related to current clinical presentation (e.g. lack of insight, current symptoms of major mental illness) and 5 items related to future risk factors (e.g. lack of personal support, non-compliance with remediation attempts). Each item was scored as 0 (not present), 1 (partially or possibly present) or 2 (present), leading to a maximum total score of 40, and maximum sub-scale scores of 20 for the historical scale and 10 for the clinical and risk scales. If insufficient information was available we omitted the item score but pro-rated the scale and sub-scales (by taking the average score on scale or sub-scale). If too many items were omitted (more than five in total, two for the historical scale and one for the clinic and risk scales), then the assessment was considered invalid and omitted from the analysis. In all we were able to score 887 patients **at their point of discharge**.

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

Comment: Describe the time point at which risk assessments were systematically performed. If the study design calls for repeated assessments at multiple time points, report the completion rate at each time point.

**R21 (Methods).** Report the number of assessors in the study as well as their training in the administration of the instrument(s) under investigation

The Static-99 was scored at the time of the 1989/90 release using only that information available at the release date. **The Static-99 was scored for each case by one investigator trained in the coding rules who had extensive experience using the Static-99 in sexually violent predator evaluations.**

Sreenivasan, S., Garrick, T., Norris, R., Cusworth-Walker, S., Weinberger, L. E., Essres, G., Turner, S., & Fain, T. (2007). Predicting the likelihood of future sexual recidivism: Pilot study findings from a California sex offender risk project and cross-validation of the Static-99. *Journal of the American Academy of Psychiatry & the Law*, 35, 454-468.

Comment: Relevant information regarding training includes formal certification through attendance of workshops and/or seminars on the administration of the instrument and/or supervision by formally trained assessors. If different instruments under investigation were administered by different assessors, specify this.

**R22 (Methods).** Identify the source(s) of information used to administer the instrument(s) under investigation

**All measures were coded from the correctional files of participants, an acceptable research method according to the measures' manuals. The PCL measures and HCR-20 recommend the use of an interview for clinical purposes, although file-based coding is acceptable and permissible for research purposes.** Further, although the HCR-20 recommends the use of a "low, moderate, high" structured final risk judgment, the manual also states that risk is generally assumed to increase with increases in the number of risk factors, making the evaluation of HCR-20 scores a necessary component of its overall evaluation (a procedure described by Douglas & Kropp, 2002). **The files are detailed and often voluminous, containing social, psychological, psychiatric, medical, criminal, and legal reports and information.**

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 & Behavior*, 32, 479-510.

Comment: Sources of information used to administer risk assessment instruments may include: criminal justice records, clinical records, school records, interviews (with evaluatee, family members, probation officers, service providers), mental health examination reports, and neuropsychological testing results.



**R23 (Methods).** Describe any modifications made to the instrument(s) under investigation

**As it was considered by the present authors to be unclear whether an offense was a noncontact sex offense or not based on the CPIC files alone and the very low frequency of official charges for noncontact sex offenses in our sample, a zero was assigned to every offender for this Static-99 item. Sex, age, number of victims, and relationship to victim (i.e., stranger, acquaintance, relative) were gathered from self-report data and used to score the corresponding Static-99 and SORAG items. Item 8 of the SORAG (“female victims under 14 years”) was scored as yes if the offender had only female victims under age 16, as this was the closest approximation of the item available for the present sample.**

Nunes, K. L., Firestone, P., Bradford, J. M., Greenberg, D. M., & Broom, I. (2002). A comparison of modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide. *Sexual Abuse: A Journal of Research & Treatment, 14*, 253-269.

Comment: Modifications made to risk assessment instruments may include: systematically adding or removing certain items or subscales, changing the scoring or weighting of items, altering cut-off thresholds or risk categories, and the use (or lack thereof) of prorating.

**R24 (Methods).** Report the geographical location and clinical setting in which risk was assessed

All subjects suspected of having committed a sexual offence, who underwent a formal forensic psychiatric evaluation (FPE) for the court at **the Department of Forensic Psychiatry, Aarhus University Hospital, or at the Clinic of Forensic Psychiatry, Ministry of Justice in Copenhagen, Denmark**, between 1 January 1978 and 31 December 1992 ( $n = 445$ ), were followed for 10.25 years as part of a larger retrospective follow-up study.

Bengtson, S. (2008). Is newer better? A cross-validation of the Static-2002 and the Risk Matrix 2000 in a Danish sample of sexual offenders. *Psychology, Crime & Law*, 14, 85-106.

Comment: Geographic location may be reported by specifying the country, region, state, city, and/or institution as appropriate. Clinical settings may include hospitals and clinics (inpatient or outpatient), correctional institutions, and/or community corrections (parole or probation offices).

**R25 (Methods).** Describe the method(s) used to recruit participants

**Participants were randomly selected from a list of all male offenders released from the two facilities during the specified period. Researchers randomly chose a starting position on the list and then selected every other file.**

Fass, T. L., Heilbrun, K., DeMatteo, D., & Fretz, R. (2008). The LSI-R and the COMPAS: Validation data on two risk-needs tools. *Criminal Justice & Behavior*, 35, 1095-1108.

Comment: Recruitment methods may include purposive selection of a subsample with specific characteristics (e.g., individuals with a substance use disorder), sequential sampling of eligible individuals during a particular time frame, convenience sampling, random selection, opt-in, opt-out, or the evaluation of total cohort. If any payments or other compensation was offered for participation, specify these.

**R26 (Methods).** Identify the temporal design of the study (prospective or quasi-prospective)

The **prospective cohort follow-up design** chosen was modelled on the MacVRAS to evaluate the predictive validity of historical, dispositional and clinical risk factors and to test the hypothesis that the non-forensic participants with high baseline scores on the VRAG, HCR-20 and VRS will be significantly more likely to be violent up to 20 weeks post-discharge than participants with low scores.

Doyle, M., Carter, S., Shaw, J., & Dolan, M. (2011). Predicting community violence from patients discharged from acute mental health units in England. *Social Psychiatry & Psychiatric Epidemiology*, 47, 627-637.

Comment: A prospective design is defined as a study design in which risk assessment instruments are administered, and participants are subsequently examined at a future time to determine whether the outcome(s) of interest occurred. A quasi-prospective design (sometimes referred to as “postdiction” or “retrospective prediction”) is defined as a study design in which risk assessment instruments are completed by a rater using information available at a past time point, and whether the outcome(s) of interest occurred is assessed at a second time point either in the future or in the past based on available archival information. If the study was quasi-prospective, state whether assessors were blind to outcome occurrence at the time when the risk assessment instrument was administered.

**R27 (Methods).** Identify the setting in which participants were followed to ascertain whether the outcome(s) of interest had occurred

**Twenty weeks after hospital discharge was chosen as the time frame for the analysis here because this was the period during which the prevalence of violence by patients in the community was at its highest (Steadman *et al*, 1998).** Research interviewers attempted two follow-up interviews with enrolled patients in the community during this period, approximately 10 weeks apart. ... Patients and collaterals independently were asked whether the patient had been involved in any of several categories of violent behaviour in the past 10 weeks (Lidz *et al*, 1993).

Monahan, J., Steadman, H. J., Appelbaum, P., Robbins, P. C., Mulvey, E. P., Silver, E., Roth, L., & Grisso, T. (2000). Developing a clinically useful actuarial tool for assessing violence risk. *British Journal of Psychiatry*, 176, 312-319.

Comment: Commonly used follow-up contexts include intra-institutional settings (e.g., jail, prison, hospital) or the community (including outpatient clinic settings). Hospital or outpatient care may be forensic (under court jurisdiction) or non-forensic. If participants were followed across multiple settings, specify each.

**R28 (Methods).** Report the average length of follow-up and time at risk (with dispersion parameter, if not fixed), including a description of periods subtracted from follow-up time (e.g., incarceration and/or hospitalization)

Inpatient violence was coded from the date of the index admission. **The average length of stay for patients in our sample was 108 days (SD=871.14 days, range=8–6366 days).** At the time of completing the HCR-20, PCL:SV, and VSC, the raters were blind to whether or not patients were violent following community release. **Depending on when they were discharged, patients were tracked in the community for an average of almost two years (M=690.26 days, SD=184.31 days, range from 312 days to 1053 days).** To increase the likelihood that comprehensive and reliable follow-up data were obtained, multiple sources were used (e.g. psychiatric hospital records, review panel office records, coroner records, additional psychiatric hospital and unit records from 16 general hospitals throughout the province with designated psychiatric units, BC Forensic Psychiatric Services records, and corrections/criminal records). Despite the limitations inherent in not using follow-up interviews with the patient and collaterals (see Steadman et al., 1998), prior research has evidenced an acceptable base rate of follow-up community violence using strictly archival methods (see Harris, Rice, & Quinsey, 1993; McNiel et al., 2003; Menzies & Webster, 1995). The University Research Ethics Review Committee at Simon Fraser University and the Ministry of the Attorney General Corrections Branch approved the research protocol.

....

To incorporate the length of time from the day of release from the hospital until the first incident of community violence, four survival analyses were carried out for men and women. Cutting scores of  $\geq 20$  on the HCR-20 total and  $\geq 8$  on the PCL:SV total were used. **Analyses for community violence took into account any time the patient spent institutionalized (e.g. hospitalized, jail).** The outcome variables for these analyses included any violence, physical violence, any crime, and violent crime following hospital discharge.

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

Comment: Length of follow-up is defined as the amount of time during which participants were observed to determine whether the outcome(s) of interest occurred. Time at risk is defined as the amount of time during which participants had the opportunity to engage in the outcome(s) of interest. Thus, in some studies (e.g., fixed follow-up studies) the length of follow-up and time at risk can be equivalent. However, time at risk can be decreased if other events intervene to reduce the participant's opportunity to engage in the outcome. For example, if the outcome is community violence, then periods of time in detention or hospital may be subtracted. Dispersion parameters, such as a standard deviation or confidence interval, should be reported to provide a measure of spread.

**R29 (Methods).** Specify the event(s) coded as meeting outcome criteria (e.g., assault, rape, homicide)

Violent recidivism was operationally defined as any criminal charge for **a violent offense against persons (assault, assault causing bodily harm, wounding, attempted homicide, homicide, kidnapping, forcible confinement, armed robbery, and all “hands-on” sexual offenses)** that occurred subsequent to the index offense. Also included were **any actions that resulted in patients being returned to maximum security (only males could qualify) that, in the judgment of the research assistants, would otherwise have resulted in such a criminal charge** (<15% of violent

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 & Human Behavior, 26*, 377-394.

Comment: If the outcome of interest was not restricted to specific events (e.g., any violent offense), state this and provide examples of events included in the definition (e.g., assault, homicide, rape, robbery). In cases where the outcome of interest was restricted to specific events, list these.

**R30 (Methods).** Identify the type (e.g., arrest, charge, conviction, incarceration) and source (e.g., criminal records, self-report, collateral) of information used to detect outcome occurrence

**Reconviction data was obtained from the Offenders Index database**, a national database covering convictions incurred in England and Wales. **Offenders were classified according to whether or not they had a sexual reconviction.** Note that those reconvicted for other kinds of offense were included in the “not sexually reconvicted”

Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research & Treatment*, 14, 139-153.

Comment: The *type* of information refers to the measurable and observable criterion that is used as an outcome. The *source* of information is defined as the documentation used to detect whether outcomes had occurred. In select cases, both the type and source of information will be the same (e.g., self-report of violence).



**R31 (Methods).** Describe the statistical methods used to conduct all analyses, and report the purpose of each analysis

**To determine the predictive validity of the revised 20-item version of the DA with its weighted scoring, the investigators used data reported from the attempted femicide cases and abused controls on the revised 20-item DA. The ability of the revised DA to correctly identify the attempted femicide cases, an independent sample, was evaluated through plots of ROC curves. Receiver operating characteristic curves represent the sensitivity and 1-specificity of a measure at each successive unit that could be a potential threshold for high risk designation. We developed estimates of the area under the ROC curve (AUC), and tested whether the AUC was greater than the chance diagonal (.500), the average value under random prediction methods. In addition, we calculated sensitivity and specificity using each of the top three levels of danger (increased danger, severe danger, and extreme danger) as a threshold for being designated high risk for attempted femicide. In addition, we also compared the mean scores on the revised 20-item DA between the three study groups (e.g., femicide, attempted femicide, and abused controls) using analysis of variance (ANOVA) and Tamhane's T2 statistic for testing whether pairwise comparisons of group means were equal when the variance differs between the groups.**

Campbell, J. C., Webster, D. W., & Glass, N. (2009). The Danger Assessment: Validation of a lethality risk assessment instrument for intimate partner femicide. *Journal of Interpersonal Violence, 24*, 653-674.

Comment: Describe statistical procedures in sufficient detail as to allow for replication of the main steps of analyses, and briefly provide a rationale for the choice of procedures.

**R32 (Methods).** Report whether risk scores and/or risk categories of the instrument(s) under investigation were used as an independent variable in analyses

The predictive validity of the violence risk assessment method was studied by logistic regression analysis. Separate analyses were performed (1) for the occurrence of any incident of violent or criminal behavior in the subsequent observational period (regardless of the occurrence of any risk enhancing behavior) and (2) for the occurrence of any risk enhancing behavior (regardless of the occurrence of any violent or criminal behavior). **The predictors studied consisted of the mean scores on the historical, clinical and situational subscales of the HKT-30, the final risk judgment by the case manager based on the HKT-30, the HoNOS-MDO mean score, and the total numbers of needs and unmet needs as assessed by the case manager on the CANFOR.**

van den Brink, R. H. S., Hooijschuur, A., van Os, T., Savenije, W., & Wiersma, D. (2010). Routine violence risk assessment in community forensic mental healthcare. *Behavioral Sciences & the Law*, 28, 396-410.

Comment: In cases where risk scores were used for some analyses and risk categories for others, specifically report what was used as the independent variable in predictive validity analyses.

**R33 (Methods).** Identify the statistical significance level used

The data were analysed by SPSS 16.0 for Windows. Intraclass correlation coefficients were used to estimate the interrater reliability for the V-RISK-10 sum score and for each item. Receiver operating characteristics (ROC) analysis was used to measure the predictive accuracy of the instrument, as ROCs are recommended for use in risk assessment studies for they are less dependent on the base rate of aggression (Mossman 1994, Douglas et al. 1999). This analysis forms a function of the true positive rate (Sensitivity) and false positive rate (Specificity). The area under the curve (AUC) displays a summary measure for the discrimination efficiency of a scale. This can range from 0.5 to 1.0 indicating the chance of a perfect ability of discrimination. One-way ANOVA was employed to analyse the possible differences between sum scores and the selected category of violence risk level assessed (low, moderate, high). The possible differences in sum scores between three groups in terms of outcome recommendation were also analysed using one-way ANOVA, and t-tests and chi-square were also used. A **conventional 5% significance level** and 95% confidence interval (CI) were employed for all analysis.

Yao, X., Li, Z., Arthur, D., Hu, L., & Cheng, G. (2011). The application of a violence risk assessment tool among psychiatric service users: A preliminary study. *Journal of Psychiatric & Mental Health Nursing*, 19, 438-445.

Comment: The statistical significance level is the threshold the investigators used to identify a likely false positive effect.

**R34 (Methods).** Describe any subgroup analyses planned *a priori*

The first analytic objective was to compare the prevalence of antisocial attitudes among the adolescent offenders to the CTS norms reported for adults in Knight et al. (2006). First, the adolescent means on each scale were compared with the adult means reported in Knight et al. (2006). Comparisons were made using an effect size statistic (Cohen's *d*) rather than a traditional test of significance. Because a significance test is heavily dependent on the specific sample size for the comparison, an index that was independent of the sample size was chosen. **Part of this objective was to determine how the study sample compared to the adult norms in terms of the variability of responses about the mean for each scale. To accomplish this objective, the percentage of the adolescent sample that was greater or less than two adult norm thresholds, the 33rd and the 67th percentiles (Knight et al., 2006), was compared.**

Dembo, R., Turner, C. W., & Jainchill, N. (2007). An assessment of criminal thinking among incarcerated youths in three states. *Criminal Justice & Behavior*, 34, 1157-1167.

Comment: *A priori* analyses are defined as those planned before the start of data analysis.

**R35 (Methods).** Report inter-rater reliability for administration of the instrument(s) under investigation (if conducted). If inter-rater reliability was not assessed, clarify why not

**Reliability of the actuarial instruments was assessed by comparing scores generated by the two independent codings. Intraclass correlation coefficients were .96 (95% confidence interval [CI] = .84, .99) for the VRAG, .95 (95% CI = .81, .99) for the RRASOR, .87 (95% CI = .57, .97) for the Static-99, and .96 (95% CI = .86, .99) for the SORAG.**

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumiere, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*, 413-25.

Comment: Commonly reported performance indicators for measuring inter-rater reliability in the forensic risk assessment literature include Cohen's kappa coefficient ( $\kappa$ ) and the intraclass correlation coefficient (ICC).

**R36 (Methods).** Include at least one discrimination performance indicator when measuring predictive validity

To test Hypothesis 1 on validity, **the predictive validity was assessed by means of ROC analysis (Mossman, 1994; Rice & Harris, 1995). We chose this statistical method because it is less reliant than other statistical analyses (like correlation coefficients) on base rates for recidivism and the particular cut-off score chosen to classify cases. Also, normality need not be assumed (Rice & Harris, 1995). ROC analyses result in a plot of the true positive rate (sensitivity) against the false positive rate (1 – specificity) for every possible cut-off score of the instrument. The resulting AUC can be interpreted as the probability that a randomly selected reoffender would score higher on the instrument than a randomly selected nonreoffender.**

Lodewijks, H. P. B., Doreleijers, T. A. H., & de Ruiter, C. (2008). SAVRY risk assessment in violent Dutch adolescents: Relation to sentencing and recidivism. *Criminal Justice & Behavior*, 35, 696-709.

Comment: The predictive validity of risk assessments can be divided into two components: discrimination and calibration. In the context of structured risk assessment, *discrimination* describes an instrument's ability to retrospectively differentiate between those who engaged in the outcome of interest and those who did not. Examples of discrimination indicators include odds ratios, correlation coefficients, and the area under the receiver operating characteristic curve. In contrast, *calibration* describes the instrument's level of fit between prospectively predicted and observed risk. Examples of calibration indicators include the positive and negative predictive values as well as the number needed to detain and number safely discharged. For a review of discrimination and calibration performance indicators, see Singh (2013).

**R37 (Results).** Report the rate of attrition

**Fifty-nine of the 390 participants could not be reached for follow-up interviews. Thus, there is complete follow-up data for 331 patients (67% of all approached) where of 34 (10%) participated only in the 10 week follow-up, 23 (7%) only in the 20 week follow-up, and 274 (83%) gave interviews on both occasions. For 83 participants (25%), additional collateral follow-up data was retrieved. A comparison of the patients with and without any follow up interview is shown in Table 1.**

Sturup, J., Kristiansson, M., & Lindqvist, P. (2011). Violent behaviour by general psychiatric patients in Sweden: Validation of Classification of Violence Risk (COVR) software. *Psychiatry Research, 188*, 161-165.

Comment: Attrition is defined as the loss of participants over the course of the study. Sources of attrition should be described, if known, including death, emigration/deportation, name changes, voluntary drop-out, inability to contact the participant and inability to obtain information from records upon follow-up.

**R38 (Results).** Report the outcome occurrence rate for the entire sample as well as for relevant subgroups

**The recidivism rate at one year post-release was 64%, with significantly more males ( $n=146$ , 70.5%) than females ( $n=41$ , 49.4%) reporting either being arrested or having committed an undetected crime ( $\chi^2 = 11.55$ ,  $p < .01$ ). The rate of violent recidivism was 17%, with significantly more males ( $n=42$ , 20%) than females ( $n=8$ , 9.6%) reporting having been arrested or having committed undetected violent criminal acts ( $\chi^2 = 4.78$ ,  $p < .05$ ).**

Hastings, M. E., Krishnan, S., Tangney, J. P., & Stuewig, J. (2011). Predictive and incremental validity of the Violence Risk Appraisal Guide scores with male and female jail inmates. *Psychological Assessment*, 23, 174-183.

Comment: Specify the number of individuals who completed follow-up (with a percentage of the total sample, for comparison) who were found to have engaged in the outcome of interest. Relevant subgroups are those that underwent subsequent analyses.



**R39 (Results).** Report predictive validity performance indicators for each outcome of interest as specified in the Methods with associated dispersion parameters

ROC analyses were performed to assess the predictive validity of the Static-99 and the SORAG. **For sexual recidivism, the Static-99 yielded an area under the curve (AUC) of .70 (SE = .05, CI = .60-.79) and the SORAG yielded an AUC of .65 (SE = .06, CI = .52-.76).** The corresponding correlation coefficients were .18 and .17 for the Static-99 and SORAG, respectively. To assess the relative predictive accuracy of the Static-99 and SORAG, their respective AUCs were compared. The difference was nonsignificant ( $Z = .97; p > .30$ ) indicating that performance was similar for both measures. ... **For violent (including sexual) recidivism, both the Static-99 and the SORAG yielded an AUC of .69 (SE = .04, CI = .60-.77).** The corresponding correlation coefficient was .23 for both the Static-99 and SORAG. As is evident from visual inspection of these results, the difference between the AUCs of the respective measures was nonsignificant ( $Z = .04; p > .90$ ).

Nunes, K. L., Firestone, P., Bradford, J. M., Greenberg, D. M., & Broom, I. (2002). A comparison of modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide. *Sexual Abuse: A Journal of Research & Treatment, 14*, 253-269.

Comment: Performance indicators statistically measure an instrument's ability to either prospectively or retrospectively assess the likelihood of an outcome of interest. Dispersion parameters, such as a standard deviation of confidence interval, provide a measure of spread.

**R40 (Results).** Report the number of participants with each risk score and/or in each risk category and how many went on to engage in the outcome(s) of interest

<b>Static-99</b>			
<b>Risk category</b>	<b>n (%)</b>	<b>Recidivists</b>	<b>Non-recidivists</b>
<b>Low</b>	<b>0 (0)</b>	<b>0</b>	<b>0</b>
<b>Medium-low</b>	<b>11 (41)</b>	<b>2</b>	<b>9</b>
<b>Medium-high</b>	<b>10 (37)</b>	<b>3</b>	<b>7</b>
<b>High</b>	<b>6 (22)</b>	<b>2</b>	<b>4</b>

Wilcox, D., Beech, A., Markall, H. F., & Blacker, J. (2009). Actuarial risk assessment and recidivism in a sample of UK intellectually disabled sexual offenders. *Sexual Aggression, 15*, 97-106.

Comment: Present the rates of the relevant outcomes by category of assessed risk. This outcome information may be displayed in a contingency table as above or in the text.

**R41 (Results).** Report the results of subgroup analyses planned *a priori* as specified in the Methods

**In the PD cohort, the VRAG predicted 2-year violent failures with an AUC of the ROC of .68 (95% CI=.62 to .73). At a cutoff of 13 points, sensitivity was .57 and specificity was .71. Positive and negative predictive values were .40 and .83, respectively. The H-10 had an AUC of the ROC of .71 (95% CI=.66 to .76). At the inflexion cutoff of 12, the sensitivity of H-10 was .72 and the specificity was .60. The positive predictive value was .38, and the negative predictive value was .86.**

**In the schizophrenia cohort, the receiver operating curve of the VRAG mounted an area of only .60, which is not significantly larger than that of the .50 line of no information (95% CI=.50 to .69). At a cutoff of 0 points, sensitivity was .68, specificity was .53, positive predictive value was .20, and negative predictive value was .91. The H-10's AUC of the ROC was .66 (95% CI=.56 to .75). At a cutoff of 8 points on the H-10, its sensitivity was .88 and its specificity was .36 (positive and negative predictive values were .19 and .95, respectively).**

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

Comment: *A priori* analyses are defined as those planned before the start of data analysis.

**R42 (Results).** Describe and report the findings of any *post hoc* analyses conducted

**Since the aim of the studied instruments is to predict future violence in violent offenders, analyses were repeated only for the respondents who committed violent offences before TBS treatment ( $n = 30$ ). Six of the seven recidivists belong to this group. Of these 30 respondents, 16 had serious addiction problems. Five of the seven recidivists belonged to the group of ex-patients with past violence combined with severe addiction. Table 4 gives an overview of the sensitivity and specificity values connected with these variables when predicting future criminal violence. Designating violent female offenders with addiction problems as possible recidivists reduces the false positive rate by a factor of two. However, as specificity rises, sensitivity is reduced, which is a general problem in risk assessment.**

Schaap, G., Lammers, S., & de Vogel, V. (2009). Risk assessment in female forensic psychiatric patients: A quasi-prospective study into the validity of the HCR-20 and PCL-R. *Journal of Forensic Psychiatry & Psychology*, 20, 354-365.

Comment: *Post hoc* analyses are defined as those not planned before the start of data analysis.

**R43 (Discussion).** Provide a summary of the principal findings, including a discussion of their relevance in the context of the current literature

**The primary goal of this research was to see if two well-established instruments for the prediction of future violence, the VRAG and the HCR-20, were applicable to a black population in the UK. Our results show that the accuracy of prediction of future violent reconvictions were good for both instruments and closely resembled the prediction efficacy for the white population.** As such we believe that this study provides a sound evidence base for the use of both of these instruments within this particular ethnic minority group discharged from medium secure psychiatric services within the UK. Of course, this does not necessarily mean that these instruments will also be effective in other ethnic minority groups, or that other instruments will also be effective in this ethnic group. Clearly further research is needed to secure a sound evidence-base for these populations. **However, the converging evidence, to which this study is supportive, appears to be that the combination of a few out of several possible risk factors can be used to produce an effective prediction instrument and that these factors appear to cross most racial and cultural boundaries.** Hence, while this direct evidence is needed, it would seem likely that these instruments will be effective in other ethnic groups.

Snowden, R. J., Gray, N. S., & Taylor, J. (2010). Risk assessment for future violence in individuals from an ethnic minority group. *International Journal of Forensic Mental Health, 9*, 118-123.

Comment: After summarizing the results of the study, the discussion section should articulate the findings with those of other studies in the literature and suggest potential implications for relevant stakeholder groups (e.g., researchers, practitioners, and policymakers).

**R44 (Discussion).** Discuss limitations of the study design

**There are several limitations to this study. First, prospective predictive research is hampered by the clinical goals of risk assessment, i.e. risk management and prevention (Dernevik et al., 2002; Hart, 1998).** Hart (1998) stated that predictions of violence are not passive assessments, but decisions that influence services delivered to individuals: “Clinicians are bound - morally, ethically, and legally - to try to prove themselves wrong when they predict violence and take every reasonable action to prevent violence” (p. 365). **In our study, clinicians were able to use the results of the HCR-20 ratings, for instance, for decisions concerning leave. Thus, it is very likely that risk management was influenced by the results of the risk assessment, for instance, high risk patients were not released from the hospital, or were separated if the risk of inpatient violence was judged to be high. So, the AUC values we obtained were already high, but might have been even higher if the results had not been used to manage risk. Second, the sample was derived from only one Dutch forensic psychiatric hospital, thereby limiting generalization.** Nevertheless, we consider this group to be representative of Dutch offenders with a tbs order, because they are largely similar in demographic, psychiatric and criminal characteristics to the total population of tbs offenders (see van Emmerik & Brouwers, 2001). **Third, the mean follow-up period of this study was somewhat limited; some patients had a very short follow-up period of only 1 or 2 months. Also, the range of follow-up periods was rather large (1-37 months), which complicates comparison between patients.** Nevertheless, the survival analyses we conducted take differences in time-at-risk into account. **Fourth, we found a rather low base rate of violence.** Although we conducted ROC analyses that are insensitive to base rates, the low base rate might have had an effect on the Cox regression analyses. **A final limitation is that data regarding violent outcome were not always reliable. Incidents of physical violence are not always reported on the information bulletins.** For example, it is possible that incidents of physical violence between patients are not observed by staff or told to staff. This is the case for inpatients, but even more so for patients who are in the transmural treatment phase or who can go outside the hospital without supervision. It should be noted, however, that most of these limitations would have had a negative effect on the predictive accuracy of the HCR-20, thus the findings might have been even stronger without these limitations.

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.

Comment: Limitations are defined as potential weakness in the study’s design and conduct that may have influenced findings.

**R45 (Discussion).** Discuss the generalizability of study findings

Whether our findings would be replicated if the numerical findings were stronger is an empirical question. **Another important question is whether the findings would be generalizable if more serious forms of violence could be measured (4). The results of our study were similar whether physical (more serious) or nonphysical (less serious) violence was used as a criterion. Similar findings have been reported for other HCR-20 studies (12).**

**Measures were coded for research purposes, so HCR-20 scores did not follow the patients. However, the treating psychiatrists probably would have included an HCR-20 completed independently for clinical practices, or a risk assessment of some kind, in their discharge summaries. It is unclear what, if any, effect this practice would have on the validity of the HCR-20 indexes collected in this study for research purposes. A higher HCR-20 score could cause increased surveillance, leading to observation of more violence. It also reasonably could lead to more effective risk management and treatment, leading to fewer episodes of violence to observe.**

Whatever the effect, it was indirect because the outpatient clinicians did not have the HCR protocols used in this study.

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.

Comment: Generalizability is defined as the extent to which the results of the study may be applicable to other populations and/or jurisdictions. Included in this is the extent to which research findings can be applied in practice.

**R46 (Discussion).** Discuss future research directions based on study findings

In spite of these limitations, the present study is one of the first evaluations of violence risk assessments completed using the revised version of the START and one of the only studies to include the START final risk estimates in the prediction models. These findings add essential new information to the growing evidence supporting START, and structured professional judgment more broadly, as approaches that clinicians can use to assess risk for a range of aggressive outcomes among adults with mental, substance use, and personality disorders. Findings also contribute to an emerging body of literature supporting the value of considering both risk and protective factors to inform assessments of violence risk. **An important next step will be to examine whether consideration of dynamic risk and protective factors, and use of START in particular, improves risk management efforts and, ultimately, reduces the prevalence and severity of aggressive outcomes. Finally, a comprehensive assessment of violence risk should include consideration of service-level (e.g., staff de-escalation tactics and training, shift change) and system-level (e.g., restraint policy) factors that may increase or decrease the likelihood of patient aggression (Gadon, Johnstone, & Cooke, 2006; Hamrin, Iennaco, & Olsen, 2009), in addition to the more “traditional” client-level assessment of violence risk (whether guided by the START or some other instrument). Although there have been some recent efforts in this area (e.g., Johnstone & Cooke, 2010), continued work is needed.**

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

Comment: After drawing out the implications and limitations of the study, the investigators should then point to the further research that is needed to clarify and further answer the questions raised by the study and its findings.



**R47 (Disclosures).** Report any commercial interests and/or source(s) of funding as well as their role(s) in the conduct of the study

**Funding: Support for this study was provided by The Swedish Research Council (Medicine). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.**

Singh, J. P., Grann, M., Lichtenstein, P., Långström, N., & Fazel, S. (2012). A novel approach to determining violence risk in schizophrenia: Developing a stepped strategy in 13,806 discharged patients. *PLoS ONE*, 7, e31727.

Comment: A commercial interest is defined as a financial conflict such as ownership of the rights to a copyrighted instrument available for purchase, or ownership of stock in a company whose value may potentially be influenced by the results of the study. Examples of roles that funders could play in the conduct of a study include salary support for co-authors and investigators; extent of control over the study design, data collection and analysis, continuation of funding or publication of the study results; access to participants (patients or inmates); permission to use an instrument or a treatment intervention such as a medication in the study.

**R48 (Disclosures).** Report whether an author or translator of the risk assessment instrument(s) under investigation was also a study author

All the data from the original MacArthur Violence Risk Assessment Study were placed on the Web at <http://macarthur.virginia.edu> in 2001. The study reported here began data collection after the original data were made publicly available. **The software was developed by COVR, Inc., in which some of the authors have a financial interest (PA, SB, TG, JM, EM, PR, LR, ES, and HS).**

Monahan, J., Steadman, H. J., Robbins, P. C., Appelbaum, P., Banks, S., Grisso, T., Heilbrun, K., Mulvey, E. P., Roth, L. & Silver, E. (2005) An actuarial model of violence risk assessment for persons with mental disorders. *Psychiatric services* 56(7): 810-815.

Comment: Disclose those study investigators who were authors or translators of the manuals and/or development studies of instruments whose predictive validity was measured.

**R49 (Disclosures).** Report whether the study presented in the article has been published in an alternative form (e.g., government report)

**Portions of the research described in this article were part of a dissertation (Langton, 2003) submitted by the first author in partial fulfillment of the requirement of the PhD degree at the University of Toronto. The dissertation research was supervised by the second author (H.E.B.). Portions of this research were presented at the 2002 annual meeting of the Association for the Treatment of Sexual Abusers (Langton, Barbaree, Harkins, Seto, & Peacock, 2002).**

Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007). Actuarial assessment of risk for reoffense among adult sex offenders: Evaluating the predictive accuracy of the Static-2002 and five other instruments. *Criminal Justice & Behavior*, *34*, 37-59.

Comment: Alternative forms of publication that warrant disclosure include book chapters, government reports, Masters theses, doctoral dissertations, and conference presentations.

**R50 (Disclosures).** Report whether the sample or a portion thereof has been studied in other publications

**The present sample included the 178 child molesters and rapists described in earlier articles (Quinsey et al., 1995; Rice, Harris, & Quinsey, 1990; Rice, Quinsey, & Harris, 1991), plus 14 sex offenders from the same samples who had been re-leased since the earlier studies. These 192 sex offenders were supplemented by 96 sex offenders from other studies of released offenders (Rice, Harris, Lang, & Bell, 1990; Rice, Harris, & Cormier, 1992).**

Rice, M. E., & Harris, G. T. (1997). Cross-validation and extension of the Violence Risk Appraisal Guide for Child Molesters and Rapists. *Law & Human Behavior, 21*, 231-241.

Comment: This item refers to the specific case in which study participants have already been included in other analyses presented in a previous journal article.

### **Closing Notes**

- 1) Information to meet checklist criteria may be provided in the text, tables, and/or figures of a manuscript. Where necessary, some criteria can be met by including information in electronic supplements that are routinely accessible to readers.
- 2) If an item is not relevant to the study (e.g., “Report the index offense composition of the sample” when the study involves a non-forensic sample), please select the “N/A” option on the checklist.
- 3) Authors of manuscripts concerning unstructured risk assessments can also benefit from the use of the RAGEE Statement. In such studies, checklist items that concern instruments should be marked “N/A”.
- 4) When RAGEE Statement reporting criteria conflict with a journal’s *Instructions for Authors*, please follow the latter.