Validation of the Risk for Sexual Violence Protocol in Adult Sexual Offenders

by

Karla J. Jackson

M.A. (Psychology), Simon Fraser University, 2008
B.A., Simon Fraser University, 1997

Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

in the
Department of Psychology
Faulty of Arts and Social Sciences

SIMON FRASER UNIVERSITY

Summer 2016

Approval

Name: **Karla J. Jackson**

Degree: **Doctor of Philosophy**

Title of Thesis: ***Validation of the Risk for Sexual Violence Protocol in Adult Sexual Offenders***

Examining Committee:

Chair: **Dr. Robert Ley**
Associate Professor

Dr. Stephen D. Hart
Senior Supervisor
Professor

Dr. Kevin S. Douglas
Supervisor
Professor

Dr. P. Randall Kropp
Supervisor
Adjunct Professor

Dr. Eric Beauregard
Internal Examiner
Professor
School of Criminology

Dr. Mario Scalora
External Examiner
Professor
Department of Psychology
University of Nebraska-Lincoln

Date Defended/Approved: August 10, 2016

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Sexual violence is a serious societal issue that is associated with victims experiencing a wide range of psychological difficulties. The proper assessment of risk for future sexual violence is critical to the treatment and management of sexual offenders. The Risk for Sexual Violence Protocol (RSVP; Hart et al., 2003) is a set of structured professional guidelines for assessing risk for sexual violence that provides a framework for estimating future risk as well as clinical formulation for treatment and management needs. To date, there has been very little research published on the RSVP even though it is currently being used by forensic professionals (Judge, Quayle, O'Rourke, Russell, & Darjee, 2014). This study examined the psychometric properties of the RSVP vis-à-vis the Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997), which is considered a parallel form of the RSVP given the similarity in content between these two sets of structured professional judgement guidelines. This study also examined the psychometric properties of the RSVP vis-à-vis a number of actuarial risk assessment instruments: the Static-99R (Helmus, Thornton, Hanson, & Babchishin, 2012), the Static-2002R (Helmus et al., 2012), and the Sexual Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, & Cormier, 1998, 2006; Harris, Rice, Quinsey, & Cormier, 2015). The sample consisted of 100 adult male sexual offenders who had participated in a sex offender treatment program. Sexual recidivism was coded over the follow-up period of approximately 10 years. Overall, the RSVP performed well in terms of interrater reliability, concurrent validity, and predictive validity. The interrater reliability of RSVP total scores and Summary Risk ratings was excellent (ICC_2 range = .85 to .96) and was comparable with the interrater reliability of the SVR-20 and actuarial instruments. The RSVP total scores and Case Prioritization ratings demonstrated good concurrent validity with respect to SVR-20 total and domain scores and with the actuarial instruments' total scores and risk categories, correlations all significant at $p < .001$. The RSVP total scores and Case Prioritization ratings were moderately and significantly associated with sexual recidivism, as were the SVR-20 total scores and the actuarial instruments' total scores and risk categories.

Keywords: Risk for Sexual Violence Protocol, risk assessment, interrater reliability, concurrent validity, predictive validity.

Acknowledgements

This dissertation would not have been possible without the never-ending support and patience of my family, particularly my amazing husband David and our two wonderful sons. I don't think that I could ever fully express the love and gratitude I have for you all.

I would like to thank Dr. Stephen Hart, my senior supervisor, for sharing his infinite knowledge, wisdom and understanding with me. I consider myself very fortunate to have had the opportunity to work with him. I would also like to thank the other members of my committee, Dr. Kevin Douglas and Dr. Randy Kropp, for providing their support and guidance with this research.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Sexual violence is an important societal issue that affects the lives of many people. Although the concept is broad in nature, a useful definition is the "actual, attempted, or threatened sexual contact with another person that is nonconsensual" (Hart et al., 2003, p. 2). It is difficult to accurately estimate the prevalence rates of sexual violence given that victims are often hesitant to either acknowledge or report incidents (Tjaden & Thoennes, 2000). Different terminology used in data collection measures can also impact on reported prevalence rates (Hamby & Koss, 2003). However, based on 2009 data, 472,000 women and 204,000 men (age 15 and older) in Canada self-reported being the victim of sexual violence in the preceding 12 months (Statistics Canada, 2010); although these figures are still likely to underestimate the scope of the problem. According to the same survey, 88% of the victims did not report the sexual violence to the police.

There are many kinds of victims of sexual violence spanning across all ages, ethnicities, socioeconomic backgrounds, and genders. Survey data indicates that up to 31% of males have been sexually abused as a child (Finkelhor, 1994; Finkelhor et al., 1986) and between 0.7 and 12.4% of adult males have been sexually victimized[1] (Coxell & King, 2010; Sorenson, Stein, Siegel, Golding, & Burnam, 1988). There is evidence to suggest that females are at greater risk for sexual victimization. For example, in a Canadian survey of 420 women, 54% reported being sexually abused as a child and 67% reported being sexually assaulted as an adult (Randall & Haskall, 1995). Governmental figures indicate that Canadian women are eleven times more likely to be sexually victimized than Canadian men (Sinha, 2013). Moreover, women who have been

[1] Depending on the population and definition used by researchers.

sexually victimized in childhood often report being revictimized in adulthood (Randall & Haskell, 1995). A review of the literature indicates that sexual victimization often leads to mental health issues such as post-traumatic stress disorder, penetration disorder, borderline personality disorder, generalized anxiety, major depressive disorder, substance abuse, and suicidal ideation or attempts (e.g., Campbell, Dworkin, & Cabral, 2009; Isely, Busse, & Isely, 1998; Sansone & Sansone, 2011; Tosh & Carson, in press; Ullman & Brecklin, 2002).

With respect to those individuals who are sexually violent, the majority tend to be male although females also sexually offend. A 1995 review of inmates under federal legislation in Canada found that of the 3,875 identified sexual offenders, 99.7% were male and 0.3% were female (Motiuk & Belcourt, 1996). Similarly, in a review of 6,838 offenders who were convicted of a sexual offence over a seven and a half year period (1985 to 1992) in Washington State, 98% were male and 2% were female (Song, Lieb, & Donnelly, 1993). However, studies on convicted populations are likely to underestimate the prevalence of female offenders. Data from a United States (US) survey that included 5,848 men, showed that 1.7% of men experienced rape in their lifetime, and of those incidents, 20.7% were committed by women (Breiding, 2015). Moreover, when a broader definition of sexual violence was used, such as being "forced to penetrate" someone, 6.7% of men (approximately 7.6 million men in the US) experienced sexual violence with 82.6% of the perpetrators being female (Breiding, 2015).

According to the 1995 Canadian review of federally legislated sexual offenders, the average age at the time of admission was approximately 38 years, with the ages ranging from 15 to 83 (Motiuk & Belcourt, 1996). There is some overlap between adult and adolescent samples of sex offenders given that the age range for adolescent sex offenders is commonly defined as age 12 to 18 (Seto & Lalumière, 2010). Within Canada, it has been estimated that adolescent sex offenders are responsible for approximately 25% of all sexual offences (Mathews, 1987). However, only a minority (between 5% to 10%) of adolescent sex offenders continue to engage in sexual violence into adulthood (Lussier, Van Den Berg, Bijleveld, & Hendriks, 2012; Nisbet, Wilson, & Smallbone, 2004; Sipe, Jensen, & Everett 1998).

Overall recidivism rates commonly range from 10% to 15% (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2009; Harris & Hanson, 2004). In a meta-analysis including 100 samples (*n* = 28,757), the sexual recidivism rate was 12% (Hanson & Morton-Bourgon, 2009). There is some evidence to suggest that a linear negative relationship exists between certain types of sexual offenders, such as rapists and incest offenders, whereby the risk for sexual violence is highest in early adulthood and then continually declines thereafter (Hanson, 2002; Hanson & Bussière, 1998). This differs from adults who commit extra-familial sexual violence against children, in that this group of offenders show little decline in sexual recidivism up until age 50 (Hanson, 2001). The sexual recidivism rates for offenders over age 50 vary significantly, with studies reporting between 4% to 35% (e.g., Hanson, 2002; Janka, Gallasch-Nemitz, Biedermann, & Dahle, 2012). Therefore, it is not surprising that there continues to be uncertainty and debate within the literature on how age impacts the risk for sexual recidivism (Lussier & Cale, 2013).

Sexual violence risk assessments are utilized in both forensic and non-forensic settings. There are numerous points at which a sexual violence risk assessment may be conducted as part of criminal legal proceedings, such as before sentencing, when determining eligibility for parole, or as part of a dangerous offender application in Canada[2] (Fabian, 2006). Risk assessments may also be conducted in non-forensic settings such as community mental health clinics and inpatient psychiatric hospitals. In these settings, sexual violence risk assessments can play an important role in the treatment and management of individuals who have a history of committing sexual violence. Risk assessments can inform decision-makers by differentiating between those offenders who require intensive treatment and those who are more likely to be managed successfully in the community (Tully, Chou, & Browne, 2013). For an assessment to effectively inform treatment and management decisions, it must provide comprehensive, accurate, and relevant information related to risk factors (Smid, Kamphuis, Wever, &

[2] The US equivalent would be an application for civil commitment of someone determined to be a sexually violent predator.

Van Beek, 2014). The assessment should also include a clinical formulation that shows how the convergence of various risk factors contributed to the individual's history of sexual violence offences, as well as how the offender's risk for future sexual violence can be potentially mitigated (Darjee & Russell, 2012).

In this thesis, I present the findings of research evaluating the interrater reliability, concurrent validity, and predictive validity of the Risk for Sexual Violence Protocol (RSVP; Hart et al., 2003). The RSVP is a set of structured professional judgement guidelines for assessing risk of sexual violence. In the remainder of the Introduction, I discuss the nature of sexual violence risk assessment and review some of the most widely used tools currently used for this purpose.

Sexual violence risk assessment

I will now provide an overview of sexual violence risk assessment and describe the context and process of its development, before examining in more detail specific sexual violence risk assessment instruments.

Overview of the development of sexual violence risk assessment

Since the 1980s, there has been a great deal of controversy in the area of sexual violence risk assessment, particularly in relation to predicting an offender's risk to sexually recidivate (Conroy & Witt, 2013). Prior to the 1950s, sexual violence risk assessments were predominantly based on the clinical judgement of the assessor (Conroy & Witt, 2013). Simply put, evaluators used their own experiences, opinions and intuition to assess an offender's risk to sexually recidivate. Using clinical judgement alone as the basis of a violence assessment has garnered a great deal of criticism given its lack of empirical support. For example, research has shown that unstructured clinical judgement performs only slightly better than chance (Grove & Meehl, 1996; Hanson & Bussière, 1998).

Meta-analyses of predictive validity indicate that clinical judgement may be less accurate or valid than other approaches to violence risk assessments[3] (Grove, Zald, Lebow, Snitz, & Nelson, 2000; Hanson & Morton-Bourgon, 2009). The previous reliance on clinical judgement, however, was out of necessity given that empirically-based risk assessment measures had not yet been developed. Widespread dissatisfaction with the over-reliance on clinical judgement alone led to a vast amount of new research, and by the end of the 1990's there were a number of empirically-supported sexual violence risk assessment measures available. The availability of new measures led to a decrease in the use of clinical judgement alone for violence risk assessments. Nevertheless, unstructured professional judgement is still relied on as the basis of some violence risk assessments (Blais & Forth, 2014).

Over time, two main approaches to assessing risk for sexual violence have become widely accepted; actuarial and structured professional judgement. Actuarial and structured professional judgement approaches have also been termed "non-discretionary" and "discretionary" respectively (Hart & Logan, 2011). The actuarial approach only allows for the inclusion of a set number of risk factors that are weighted and combined based on an algorithm to provide an estimate of risk for future violence (Pedersen, Rasmussen, & Elsass, 2010). To be included, the risk factors need to have empirical, theoretical, or practical support. Actuarial violence risk assessments have been developed for the purpose of predicting violence recidivism within a certain population and within a specific timeframe (Hart & Logan, 2011). For example, an actuarial measure may provide an estimate of risk to sexually recidivate among a population of known sexual offenders within the next 5 years. The structured professional judgement approach provides evaluators with a set of guidelines for decision-making regarding the assessment of risk for future violence. These guidelines are based on the breadth of forensic psychology, including both empirical data and

[3] For example, see actuarial and structured professional judgement approaches described later in this chapter.

professional practice. The structured professional judgement approach is often referred to as guided clinical judgement.

A key difference between actuarial and structured professional judgement approaches and clinical judgement alone is that the former two approaches are generally considered to be evidence-based whereas the latter lacks empirical support. The standard of using an evidence-based approach has been adopted by the discipline of violence risk assessment (including sexual violence). The concept of evidence-based approaches originated within the field of medicine. It posits that patient care should be based on the critical appraisal and application of the best available evidence, which includes both individual clinical expertise and empirically-supported clinical research and evidence (Sackett, Rosenberg, Muir Gray, Haynes, & Richardson, 1996). However, there are divergent perspectives regarding how to define evidence-based decision-making, its principles, and how it should be carried out in practice (Hart & Logan, 2011). These differences are associated with either a “latitudinarian” or “orthodox” position (Hart, 2003a). The latitudinarian position considers the evidence-based approach to be an underlying principle that guides decision-making, allowing for clinical discretion concerning what information is considered and how it is weighted. Clinical discretion is viewed as necessary given that empirical and professional literature is accepted as inherently incomplete and imperfect. The orthodox perspective on evidence-based decision making is more rigid, favoring a structured, algorithmic approach derived directly from empirical research. It avoids or minimizes clinical discretion under the assumption that human judgement is untrustworthy (Hart, 2003a).

Although both structured professional judgement and actuarial approaches are viewed as evidence-based by their supporters, each approach has been developed from a different perspective. The structured professional judgement approach embodies the latitudinarian perspective. This approach is informed by a large body of scientific and professional literature. There is also considerable empirical support for the reliability and validity of violence risk assessments based on structured professional judgement guidelines (e.g., Otto & Douglas, 2010). The actuarial approach is founded on the orthodox or narrower perspective and has empirical support regarding its reliability and validity (Quinsey et al., 2006). However, the extent to which it is evidence-based is

potentially undermined by the limited number of items included in actuarial measures, and the exclusion of other potentially relevant items or professional knowledge.

Current perspectives on sexual violence risk assessment

Within the field of forensic psychology there has been an ongoing debate regarding whether actuarial measures or structured professional guidelines are more appropriate for sexual violence risk assessment (e.g., Conroy & Witt, 2013; Craig, Beech, & Harkins, 2009; Hanson & Morton-Bourgon, 2009). Hart and Logan (2011) provide a useful framework for evaluating and comparing these two approaches that includes looking at the efficacy, efficiency, and utility of an approach. Within this framework, efficacy refers to a violence risk assessment approach that has been found to be reliable and valid in controlled research settings. Efficiency speaks to the consistency of an approach to violence risk assessment as being reliable and having predictive validity within real-world forensic settings. Within this framework, utility is defined as the extent to which users or evaluators perceive the approach to be useful and socially valid.

There is a body of empirical data that supports the efficacy of both actuarial and structured professional judgement approaches to violence risk assessment. Overall, both approaches have been found to have good to excellent interrater reliability within controlled research studies, with actuarial measures having slightly better interrater reliability on average (Tully et al., 2013). Similarly, both actuarial and structured professional judgement approaches tend to have fair to moderate predictive validity (Guy, 2010; Hanson & Morton-Bourgon, 2009; Singh, Grann & Fazel ,2011; Tully et al., 2013). Within research, there is variability regarding which approach performs better, although in general, the difference between these two approaches is relatively minor.

In terms of efficiency, given the complexity of violence risk assessments, it can be challenging to conduct research looking at the reliability and validity of assessment approaches in field studies. Nevertheless, there has been research to support the reliability and validity of both actuarial and structured professional judgement approaches to violence risk assessment within field settings. Otto and Douglas (2010) provide a comprehensive summary of the violence risk research including field data on

actuarial and structured professional judgement approaches. On average, both approaches have good interrater reliability and fair to moderate predictive validity within field studies (Hart & Logan, 2011).

The utility of an approach to violence risk assessment is as important as the efficacy and efficiency, although this area is only beginning to receive attention with a relatively small number of studies having been conducted. Within the literature, both actuarial and structured professional judgement approaches are viewed to be useful and socially valid[4] (e.g., Campbell, French, & Gendreau, 2009). However, there is some evidence to suggest that the structured professional judgement approach is viewed more favorably in terms of utility. Within the context of decision-making for risk management, stakeholders tend to prefer the structured professional judgement approach (Monahan et al., 2001; Pedersen et al., 2010). Relative to an actuarial approach, forensic professionals have rated a structured professional judgement approach more positively because of its inclusion of dynamic factors and ability to inform treatment (Khiroya, Weaver, & Maden, 2009). An important shift in violence risk assessment related to utility has been the increasing awareness of the need for assessments to include a comprehensive clinical formulation. As noted earlier, clinical formulation provides an understanding of how an offender's risk factors contribute to acts of violence and helps to target treatment and management strategies so as to lessen the likelihood of future violence. The actuarial approach to violence risk assessment has been criticized for not providing a suitable framework for comprehensive clinical formulation (Hart & Logan, 2011) and for being unable to integrate changes in dynamic factors that may affect risk (Tully et al., 2013).

To summarize, actuarial and structured professional judgement approaches are considered to have efficacy and efficiency. Both approaches have been viewed as having utility, with the structured professional judgement approach offering additional

[4] Some, however, question the usefulness and validity of using these approaches with certain populations, particularly in relation to gender, race, and disability (e.g., Davidson & Chesney-Lind, 2009; Hannah-Moffat, 2013; Zinger, 2004)

advantages regarding the treatment and management of violent offenders. Consequently, forensic professionals use both approaches. Blais and Forth (2014) reviewed 111 risk assessment reports within Canada that had been submitted to court for the purpose of informing either Dangerous Offender or Long-Term Offender applications. Overall, actuarial measures were included in 92% of the reports and structured professional judgement measures were used in 53% of the reports, suggesting that a significant number of the reports included both kinds of measures. In a survey of forensic psychologists, the two most widely used sexual violence risk measures were the Static-99 (Hanson & Thornton, 1999) and the Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997), an actuarial and a structured professional judgement measure respectively. Of these two instruments, the Static-99 was used more frequently (Archer, Buffington-Vollum, Stredny, & Handel, 2006).

Having provided a brief background on the development of sexual violence risk assessment, I will now examine specific instruments currently used in a range of forensic and non-forensic settings. These include, the Static-99, the Static-99R, the SVR-20, and the RSVP. These instruments were selected due to their frequent use and popularity within the profession. For each instrument, I discuss the interrater reliability, predictive validity and utility when possible.

The Static-99

The Static-99 (Hanson & Thornton, 1999) is an actuarial risk assessment instrument developed for the purpose of estimating risk for sexual and violent recidivism among adult male offenders who have committed at least one act of sexual violence. The Static-99 features 10 items of a static or historical nature that have empirical support for their relationship to future sexual and non-sexual violence. According to Hanson and Thornton (1999), these 10 items are categorized into five types of risk factors: sexual deviance (male victims, never married and non-contact sex offences); range of potential victims (unrelated victims and stranger victims); persistence (prior sex offences); antisocial (current non-sexual violence, prior non-sexual violence and four or more sentencing dates); and age (18-25.99 years). Each item is scored either 0 or 1, with the exception of one item (prior sex offences) whose score can range from 0 to 3; total

scores range from 0 to 12. Total scores are translated into a risk category: 0 or 1 = *low*, 2 or 3 = *moderate-low*, 4 or 5 = *moderate-high*, and $\geq$ 6 = *high*. The risk categories were created based on survival analysis to estimate the sexual recidivism risk associated with total scores (Sreenivasan, Weinberger, Frances, & Cusworth-Walker, 2010). In terms of recidivism rates associated with Static-99 total scores, it has been recommended that relative risk (e.g., one offender's risk to reoffend compared to other offenders' risk) be reported as opposed to absolute risk (e.g., based on this offender's Static-99 score, he has a 47% chance of reoffending) (Hanson, Babchishin, Helmus & Thornton, 2013). According to Hanson et al. (2013), relative risk is best reported in the form of a risk ratio. For example, each increase in total score is associated with a 20% increase in the "relative" likelihood of reoffending, as quantified by odds ratios or rate ratios of approximately 1:2.

Static-99 Interrater Reliability

Numerous studies have examined the interrater reliability of the Static-99 (e.g., Bengtson, 2008; de Vogel, de Ruiter, van Beek, & Mead, 2004; Olver, Wong, Nicholaichuk, & Gordon, 2007; Rettenberger, Matthes, Boer, & Eher, 2010; Sjöstedt & Långström, 2001). Overall, the Static-99 is reported to have high interrater reliability. Hanson (2001) compared the Static-99 ratings for 55 individuals involved in civil commitment hearings in the US. The ratings were completed by independent evaluators and were found to have an average item rating agreement of 91%, an item kappa of .80, and total score intraclass correlation (ICC)[5] of .87. Looman (2006) and Harris, Phenix, Hanson, and Thornton (2003) found similar results for the Static-99 within a Canadian

[5] ICCs actually comprise a family of cofficients, each of which is intended to be used in different situations. Unfortunately, it is often the case that stuies do not specify which type (model, form) of ICC was calculated. Throughout this thesis, I will report the specific type of ICC calculated in a study using the authors' notations, where available; I will use ICC_1 to refer to Model 1 ICC for single ratings, calculated for absolute agreement using a mixed effects model; I will use ICC_2 to refer to Model 1 ICC for averaged ratings, calculated for absolute agreement using a mixed effects model; and I will simply use ICC when it is unclear what model or form of ICC was calculated. Regardless of model or form, ICC can take values from -1.0 to 1.0, and Fleiss (1981) recommended the following interpretive guidelines: < .39 = poor, .40 to .49 = fair, .50 to .74 = good, and > .75 = excellent.

sample of high-risk federal offenders, with Looman reporting a Pearson correlation of .90 (*n* = 26) and Harris et al. reporting an ICC of .87 (*n* = 10). Using a larger sample of Canadian offenders, Hanson, Harris, Scott, and Helmus (2007) compared the Static-99 ratings for offenders on community supervision for a sexual offence. The ratings were completed by trained probation officers and the interrater reliability was excellent, ICC = .91 (*n* = 92). In a recent review of the empirical literature pertaining to sexual violence risk assessment measures, the interrater reliability for the Static-99 was found to be very good overall (Tully et al., 2013). The studies included in this review reported interrater reliability using different analyses, which were Pearson's correlations (*r* ranging from .88 to .90), average Cohen's Kappa (κ = .90 for both studies), and intraclass correlations (ICC_1 ranging from .80 to .94). In contrast, a study done by Ducro and Pham (2006) found that the interrater reliability for the Static-99 was much lower (ICC = .63), despite the raters being psychologists (MA-level) who closely followed the Static-99 manual.

A great deal of the research on the Static-99 interrater reliability has been conducted under controlled conditions and included specific Static-99 training for the raters. In response to this, Quesada, Calkins, and Jeglic (2014) compared Static-99 ratings done by practitioners in the field and trained researchers. The sample consisted of 1,973 adult sexual offenders who had been released from a state correctional facility. On an individual item level, the consistency between practitioners and researchers was indexed using κ and ranged from .62 to .94. In approximately 60% of the cases, there was no discrepancy between practitioner and researcher rating. In the approximately 40% of cases where there was a discrepancy, the total scores fell into a different risk category in over half of the cases. Researchers were associated with slightly higher total scores than the practitioners, although both groups of raters were equivalent in terms of whether the total score led to a higher risk category. The interrater reliability of the researchers was evaluated in a sub-sample of 30 cases and ranged from fair to excellent (total scores, ICC = .89; individual items, κ ranging from .27 to .87). The reliability of Static-99 within a field context was also examined by Rice, Boccaccini, Harris, and Hawes (2014). These researchers analysed the Static-99 ratings of 23,684 male offenders, all of whom had been evaluated with the Static-99 at least twice over the same period of incarceration. Although the Static-99 total score interrater reliability was good, ICC_{A1} = .81, 95% CI [.81, .82], the reliability of the scores decreased as the total

score increased. The odds of there being disagreement between the ratings when the initial Static-99 score was 6, was approximately 3.16 times than if there being disagreement with an initial score of 1. The authors highlight the potential implications of this finding as the Static-99 may be used to identify high risk sexual offenders for the purpose of civil commitment or highly restrictive supervision orders.

Static-99 Predictive Validity

The predictive validity of the Static-99 has been widely researched across a number of different countries, including Canada, the US, and the United Kingdom (UK). It is commonly evaluated based on the area under the receiver operating characteristic curve (AUC) (see Rice & Harris, 2005). Based on a review of the literature, Anderson and Hanson (2010) concluded that the Static-99 has been found to have an AUC of approximately .70, which suggests that the Static-99 is a moderate predictor of sexual recidivism. Since that time, there have been several studies that lend support to the predictive validity of the Static-99. In a metaregression analysis done by Singh et al. (2011), the Static-99 had moderate predictive validity (*Mdn* AUC = .70; IQR = .62 to .72) based on 12 studies (*n* = 8,246). Similarly, in a review of research examining the predictive accuracy of sexual violence assessment measures, Tully et al. (2013) applied specific quality assurance parameters that led to the inclusion of 30 studies (*N* = 20,727) related to the predictive validity of the Static-99. Based on these studies, the Static-99's mean AUC was found to be .69 (range = .57 to .92). Moreover, in a recent study by Smid et al. (2014) it was found that the AUC for the Static-99 was .72 at the 5-year follow up and .73 at the 10-year follow up. The Static-99 was coded based on the file review of 397 male adult sexual offenders in the Netherlands.

However, there is variability across the research, with some studies finding that the Static-99 did not predict sexual recidivism within certain ethnicities. For example, Långström (2004) examined the predictive accuracy of the Static-99 in a group of 1,400 adult male sexual offenders who had been released from prison. Although the Static-99 moderately predicted sexual recidivism for Nordic and European offenders, it did not predict sexual recidivism in African or Asian offenders. Some studies have also found that it is a better predictor of sexual recidivism for certain types of sexual offenders. For example, in a study by Bartosh, Garby, Lewis, and Gray (2003), the Static-99

significantly predicted sexual recidivism for offenders who committed child molestation (incest and extra-familial) but not for rapists or non-contact offenders. However, when the total sample was considered (*n* = 186), the Static-99 was found to significantly predict sexual recidivism. Similarly, Harris et al. (2003) found that the Static-99 was a better predictor of sexual recidivism for offenders who commit child molestation than those who rape.

The Static-99R

In response to increasing empirical evidence that sexual recidivism rates decline as sex offenders get older (especially from age 60 onwards), the Static-99 was revised to more accurately assess aging as a risk factor (Helmus, Thornton, Hanson, & Babchishin, 2012). Hanson, Harris, Helmus and Thornton (2014) found that recidivism risk was highest for high risk offenders in the first few years following their release. The longer they remained in the community and offence-free, the greater the decline in recidivism risk.[6] In the Static-99, age is coded based on the offender's current age and is dichotomized; aged 25 and over (coded 0) and aged 18 – 24.9 (coded 1). As noted earlier, Static-99 total scores can range from 0-12 and the total scores are associated with risk categories. In the Static-99R, the offender's age at release is coded as follows: 18 to 34.9 = 1; 35 to 39.9 = 0; 40 to 59.9 = -1; and 60 or older = -3. Static-99R total scores can range from -3 to 12. The risk categories for the Static-99R were established using logistic regression as opposed to survival analysis, which was used to create the Static-99 risk categories. The Static-99R also provided new normative data with the sample being divided into normative subgroups, such as "routine correctional samples" and "those pre-selected for treatment needs," among others. For each subgroup, 5 and 10 year recidivism rates are provided for total scores, as well as the overall relative risk ratios associated with total scores (Phenix, Helmus, & Hanson, 2015). There have also been advances in the development of percentile ranks for the Static-99R, although the

[6] However, it was also noted that evaluation based on subsequent offenses can be problematic as many offenses are undetected and unreported (Hanson et al., 2014).

authors caution that such measurements should not be used in isolation to assess overall offender risk (Hanson, Lloyd, Helmus & Thornton, 2012).

Static-99R Interrater Reliability

Due to the recent revision of the Static-99, there are currently very few studies available that analyse the interrater reliability of the new version. These limited initial studies, however, report excellent interrater reliability. For example, McGrath, Lasher and Cumming (2012) found excellent interrater reliability (ICC = .89) in a study where two raters coded 109 Static-99R for adult male sex offenders. Another study, by Hanson, Lunetta, Phenix, Neeley and Epperson (2014), involved 55 participants that were parole officers, probation officers and clinicians who had been trained in the use of the Static-99R (but had varying levels of experience). Each rater coded 14 files and the interrater reliability for the Static-99R total scores was calculated using the intraclass correlation coefficient (ICC) for absolute agreement using a two-way, random effects model. Overall, interrater reliability for total scores was excellent, ICC = .78, 95% CI [.64,.90]. The Pearson correlations between the "correct" scores provided by the certified trainer and the average scores provided by the raters were also high (r = .90 to .97). On average, approximately half of the scorers gave the correct score, but most of the raters were within 1 point of the correct score for most of the cases. There was a meaningful difference, however, between the experienced[7] coders (ICC = .85, n = 33) and the less experienced coders (ICC = .71, n = 22).

Static-99R Predictive Validity

As with interrater reliability studies, predictive validity studies are still emerging in response to the revised Static-99 assessment tool. Overall, the findings of these initial studies show the AUC for the Static-99R to be between .69 and .82. For example, Helmus, Hanson, Thornton, Babchishin and Harris (2012) conducted a meta-analysis of

[7] 'Experienced' was defined as having coded 26 or more real cases prior to participating in the study.

23 studies, which included 8,106 participants. For the combined data set, the Static-99R AUC was .71. When the AUCs from each study were analysed, the average weighted AUC (fixed-effect) was .69, and the random effects AUC was .70. The authors concluded that for every increase of one point on the Static-99R total score, there was an increase in the odds of sexual recidivism of 1.34. The predicted recidivism rates for both Static-99R and Static-2002R demonstrated large and significant variability across studies. These findings are consistent with a meta-analysis conducted by Babchishin, Hanson, and Helmus (2012), which included 20 samples (*n* = 7,491). They used fixed-effect and random-effects meta-analyses to compute the AUC and concluded that the Static-99R AUC was approximately .69.

Individual studies have reported slightly higher AUC scores. A study by Hanson et al. (2014) included 475 adult male sex offenders who had been released from a correctional facility in 2006-2007. The AUC for both the Static-99 and Static-99R was .82 at the 5-year follow-up. Other studies have reported slightly lower AUCs, such as Smid et al. (2014) who found that for sexual recidivism, the AUC for the Static-99R was .74 at both the 5 and 10 year follow-up. Similarly, Brouillette-Alarie and Proulx's (2013) research with 711 male sex offenders in maximum security facilities found the overall AUC for the Static-99 to be .72, and for the Static-99R to be .73. For those who had been convicted of sexually assaulting children, the Static-99 and the Static-99R AUC was .77. The assessment of offenders who had sexually assaulted women resulted in a Static-99 AUC of .70 and a Static-99R AUC of .73 (Brouillette-Alarie & Proulx, 2013). Moreover, Hanson et al. (2013) examined the relative risk associated with Static-99R scores using eight samples of adjudicated sex offenders (*n* = 4,037). Based on Cox regression calculated with continuous Static-99R scores, the average hazard ratio was 1.39, 95%CI [1.33, 1.46]. The variability of risk ratios across samples was evaluated and reported as being no greater than what would be expected by chance and there was little study variance.

Studies that have examined the predictive validity of the Static-99R on different ethnic groups have found lower AUC scores. For Indigenous sex offenders in Australia, Spiranovic (2012) argued that while the Static-99 and the Static-99R are acceptable in predicting violent recidivism, neither assessment tool significantly predicted sexual recidivism in Indigenous offender populations. This is supported by Smallbone and

Rallings' (2013) study that involved 399 adult sex offenders. They found that the predictive validity for Indigenous offenders using the Static-99R was marginal for any violent recidivism, with an AUC of .65, and that it did not predict sexual recidivism with this population of offenders, with an AUC of .61. Therefore, further research is needed regarding the validity of the Static-99R on culturally and ethnically diverse populations.

To summarize, the Static-99 and Static-99R are widely used actuarial risk assessment instruments that have generally been found to have good interrater reliability and predictive validity. I will now review the literature regarding the SVR-20, a structured professional judgement approach to sexual violence risk assessment, specifically with regards to interrater reliability and predictive validity.

The Sexual Violence Risk-20 (SVR-20)

The intention behind the SVR-20 (Boer et al., 1997) was to provide a wide array of criminal justice professionals with the means to conduct a thorough sexual violence risk assessment that was empirically-based and sensitive to dynamic risk factors. As noted earlier, the SVR-20 is reported to be the second most widely used sexual violence risk assessment measure by forensic psychologists (Archer et al., 2006). The SVR-20 is made up of 20 individual risk factors from four domains: Psychological adjustment, Social adjustment, History of sexual offences, and Manageability. Each risk factor is rated using a 3-point ordinal scale (*Absent*, *Possibly/partially*, *Present*). A summary risk rating (*low*, *moderate*, or *high*) is made based on the integrative judgements about the risks posed by the individual in light of the presence of the individual risk factors.

SVR-20 Interrater Reliability

Sjöstedt and Långström (2002) found the SVR-20 individual items to have poor to fair interrater reliability (average Cohen's $\kappa = .36$), which was suspected to be related to inadequate training of the raters. After additional training, the interrater reliability improved (average Cohen's $\kappa = .51$, *Mdn* = .57, range = .08 to 1.00). The SVR-20 summary risk ratings had fair interrater reliability (Cohen's $\kappa = .50$). A subsequent study found that the interrater reliability for the SVR-20 individual items was fair or better for 18 of the individual items (de Vogel et al., 2004). The other two items, Sexual Deviance

(ICC_1 = .38) and Relationship Problems (ICC_1 = .29), were thought to have low interrater reliability due to the lack of clinical experience of one of the raters and a lack of variance respectively. The SVR-20 summary risk ratings had fair interrater reliability (ICC_1 = .48). However, the section scores had good to excellent interrater reliability, with an ICC_1 range of .74 to .78. The interrater reliability of the SVR-20 individual risk items was also evaluated by Ramírez, Illescas, García, Forero, and Pueyo (2008) and ranged from good to excellent (average Cohen's κ = .95, range = .73 to 1.00). Zanatta (2006) rated the SVR-20's individual risk factors and recoded them numerically to create two summary scores (Psychological adjustment and Sexual offence), and found excellent interrater reliability for both summary scores (ICC_1 = .87). Several studies have looked at the interrater reliability of the SVR-20 total scores, created by recoding the individual items numerically and adding them (Hill, Haberman, Klusmann, Berner, & Briken, 2008; Rettenberger, Boer, & Eher, 2011; Rettenberger & Eher, 2007; Smid et al., 2014). Overall, these studies reported excellent interrater reliability for SVR-20 total scores (ICC_1 range = .84 to .87). One study (Barbaree, Langton, Blanchard, & Boer, 2008) reported "moderate-high" interrater reliability (Spearman's Rho = .75).

SVR-20 Predictive Validity

In a study conducted by Dempster (1998), the SVR-20 and other sexual violence risk assessment instruments were coded based on file information for 95 adult male sex offenders in Canada. The SVR-20 individual items were recoded and calculated to create total scores, which were evaluated along with the summary risk ratings. Both the total scores (AUC = .74, $p < .001$) and summary ratings (AUC = .77, $p < .001$) were found to discriminate between sexually violent recidivists and nonrecidivists. Summary risk ratings differentiated between sexually violent recidivists and nonsexually violent recidivists (AUC = .68, $p < .05$), however, total scores did not (AUC = .55, n.s.). Summary risk ratings also had unique predictive power with respect to recidivism when controlling for SVR-20 total scores using incremental validity analyses. When compared to other sexual violence risk assessment instruments, the predictive validity of the SVR-20 was either equal to or better than other risk assessment measures.

In contrast, Sjöstedt and Långström (2002) found that according to ROC analyses, the SVR-20 summary risk ratings and total scores did not significantly predict

recidivism in a sample (*n* = 51) of adult male sexual offenders (all AUC $\leq$ .56). The study also looked at other risk assessment measures and found that none of the measures predicted recidivism based on correlational analyses. The interrater reliability for this study was somewhat limited, which may have contributed to the lack of significant results, particularly as a positive correlation has been found between interrater reliability and predictive accuracy (Hanson & Morton-Bourgon, 2009; Smid et al., 2014).

Using a prospective longitudinal research design, Rettenberger et al. (2011) evaluated the predictive accuracy of the SVR-20 using a sample of 493 sexual offenders released from the Austrian prison system. The SVR-20 total score had good predictive accuracy for sexual recidivism within the total group of offenders (AUC = .72). The SVR-20 total score was also found to significantly predict sexual recidivism for both rapists (*n* = 221, AUC = .71) and child molesters (*n* = 249, AUC = .77). The authors concluded that, "the predictive accuracy of the SVR-20 can be regarded as at least as good as the predictive accuracy of other risk assessment instruments such as, for instance, the Static-99" (p. 1020).

A number of studies have examined the predictive accuracy of the SVR-20 and Static-99 with mixed results. In a recent meta-analysis done by Hanson and Morton-Bourgon (2009), SVR-20 final risk ratings (*low*, *moderate*, or *high*) were found to be the best predictor of sexual violence recidivism (average *d* = 1.11), however, this was based on a small number of studies. Comparatively, the actuarial measures included in the meta-analysis had effect sizes that ranged from *d* = .67 to .97, with the Static-99 falling at the low end of that range (average *d* = .67). In some studies, the SVR-20 has been found to have better predictive accuracy than the Static-99 with regards to sexual recidivism (Singh et al., 2011; Stadtland et al., 2005; Tully et al., 2013). For example, de Vogel et al. (2004) looked at the predictive accuracy of the SVR-20 and Static-99 in a study with 122 adult male sexual offenders in the Netherlands. The SVR-20 and the Static-99 were coded using file information. Based on ROC analyses, the SVR-20 total scores and summary risk ratings had good predictive validity (AUC = .80 and .83 respectively). The Static-99 total scores were also predictive, although less so (AUC = .71). In addition, incremental validity analyses supported that SVR-20 summary risk ratings had some unique predictive power with respect to recidivism even when SVR-20 numerical scores were controlled for.

The results of other studies indicate that the Static-99 is a stronger predictor of sexual recidivism than the SVR-20 (Parent, Guay, & Knight, 2011; Smid et al., 2014). Some studies have found that neither the SVR-20 nor Static-99 were significantly associated with sexual recidivism (Craig, Browne, & Beech, 2006; Hill et al., 2008; Rettenberger et al., 2010). It should be noted that in the above studies, the SVR-20 and Static-99 often performed similarly with the differences between their predictive accuracy being non-significant. Furthermore, according to the structured professional judgement approach, "predictive validity" is but a single, flawed way to evaluate a risk assessment tools (Hart & Logan, 2011).

The Risk for Sexual Violence Protocol (RSVP)

Despite the popularity and utility of the SVR-20, it is limited as it does not provide users with guidelines for developing strategies for managing a sex offender's risk for recidivism. Recognizing the need for sexual violence risk assessments to go beyond providing a final risk rating and predicting risk to recidivate, the RSVP was developed (Hart et al., 2003). The RSVP is a set of structured professional judgement guidelines designed to facilitate comprehensive sexual violence risk assessments that are both treatment and management oriented. The RSVP was developed to expand on the SVR-20. Although the RSVP and SVR-20 have very similar content, the RSVP has a stronger focus on psychological risk factors. It also includes steps for identifying likely scenarios for future sexual violence and treatment or management strategies that would help to effectively ameliorate future risk for sexual violence. It is suggested in the RSVP manual that the identified scenarios speak to the nature, severity, imminence and likelihood of future sexual violence. It is also suggested that evaluators provide a Case Prioritization rating (as opposed to an overall risk rating), which reflects the level of treatment and/or management required to lessen the offender's risk for future sexual violence (Hart et al., 2003). The RSVP has become a widely utilized sexual violence risk assessment instrument (Hart & Boer, 2010), yet to date, there has been very little published research on the reliability and validity of the RSVP.

RSVP Interrater Reliability

The interrater reliability of RSVP ratings was assessed by Hart (2003b) in a Canadian sample of 50 low-risk adult male sex offenders who were in the community on supervision. The RSVP ratings were completed by two independent raters who coded file information only. For the individual risk factors, the interrater reliability was good to excellent and ranged from ICC_1 = .58 to .97 (*Mdn* = .91) for past ratings, ICC_1 = .62 to 1.0 (*Mdn* = .87) for recent ratings, and ICC_1 = .65 to .95 (*Mdn* = .88) for future relevance ratings. The interrater reliability for domain scores (Sexual violence history, Psychological adjustment, Mental disorder, Social adjustment, Manageability, and overall total scores) was excellent. The past domain total score ratings were as follows: ICC_1 = .93 for Sexual violence history, .92 for Psychological adjustment, .96 for Mental disorder, .96 for Social adjustment, .98 for Manageability, and .99 for total overall score. The recent domain total score ratings were as follows: ICC_1 = .93 for Sexual violence history, .90 for Psychological adjustment, .96 for Mental disorder, .87 for Social adjustment, .90 for Manageability, and .96 for overall total score. The relevance domain total score ratings were as follows: ICC_1 = .93 for Sexual violence history, .90 for Psychological adjustment, .95 for Mental disorder, .85 for Social adjustment, .93 for Manageability, and .98 for overall total score. The interrater reliability for the Case Prioritization was ICC_1 = .68.

In a subsequent study, Watt, Hart, Wilson, Guy, and Douglas (2006) evaluated the interrater reliability of RSVP ratings in a sample of 50 high-risk adult male sex offenders who were on community supervision. Two independent raters completed the RSVP ratings based on file review and made consensus ratings for each examinee. The results were consistent with the Hart (2003b) study. In the Watt et al. (2006) study, the interrater reliability for individual risk factors was good to excellent and ranged from ICC_1 = .58 to 1.0. The interrater reliability for summary scores was excellent and ranged from ICC_1 = .88 to .99. The interrater reliability for Case Prioritization was also excellent, ICC_1 = .92.

In an effort to determine the extent to which training and experience may influence the interrater reliability of RSVP ratings, Sutherland et al. (2012) conducted a study involving 28 clinicians with varying levels of experience and training. These

clinicians coded the RSVP based on six case vignettes and their level of agreement was evaluated. The clinicians' ratings were also compared to "gold-standard" ratings that had been completed by forensic evaluators who were highly trained in conducting sexual violence risk assessments. Sutherland reported that overall, the interrater reliability on an item level was fair, ranging from poor to excellent, ICC_2 = .05 to .78, *M* = .51. The interrater reliability for the domain and summary ratings was fair to good, ranging from ICC_2 = .45 to .52 and from ICC_2 = .43 to .69, respectively. The interrater reliability for Case Prioritization was also good, ICC_2 = .62. Sutherland et al. (2012) found that interrater reliability increased in relation to the clinicians' amount of training, particularly with regards to the RSVP.

A recent study done by Wilson (2013) examined the RSVP with respect to interrater reliability. This study involved 17 professionals completing an online RSVP training program and then completing RSVP ratings for six of ten possible cases. The interrater reliability analyses considered ratings across six raters for each case and were analysed using Type 1 intraclass correlations for single raters and absolute agreement. The interrater reliability for individual items was generally poor to fair (19 out of 44 items had ICCs below .40, 17 out of 44 had ICCs that ranged from.40 to .59), with four items having good agreement (ICCs ranged from .60 to .74) and one item having excellent agreement (ICC = .91). The RSVP total scores had fair rater agreement (Presence total scores ICC = 56; Relevance total scores ICC = .55). The RSVP domain scores were associated with fair to excellent interrater reliability, ICCs ranged from .43 to .78, with the exception of the Sexual Violence History domain scores, which had poor rater agreement (Presence ICC = .07; Relevance ICC = .12). For each of the cases that were assessed by the participants, gold standard ratings had been completed by experts. The participants' ratings were compared with the gold standard ratings to determine the "percent agreement" for individual items and summary risk judgements. The average item agreement was higher for Presence ratings (76%) than for Relevance items (64%). The percent agreement for the summary risk ratings were lower than for individual items: Case Prioritization 59%; Risk for Serious Harm 59%; and Immediate Action Required 38%. In addition, the gold standard total scores were significantly higher than the participants' total scores, based on t-tests.

Finally, Darjee et al. (2016) examined the interrater reliability of the RSVP. They studied 109 individuals who had been assessed by the Sex Offender Liaison Services (SOLS) in Edinburgh, Scotland. Although the majority of those included in the study had a previous conviction, a minority had no conviction but had been assessed to be at risk of committing sexual violence, such as having sexually violent fantasies. Moreover, some convictions were not related to sexual violence directly (i.e., murder or assault), but the crimes were deemed to have a sexual component. The authors concluded that the sample was a particularly high risk and complex group, and therefore was not representative of more general forensic samples or suitable for comparison where research used case reports of offenders with only a prior conviction. The assessments were completed by two clinicians between 2006 and 2013, in collaboration with a multidisciplinary team. Interrater reliability was based on the coding of 11 of the cases being independently rated by two evaluators. Interrater reliability was analysed using intraclass correlations for both one rater (ICC_1) and the average reliability across both raters (ICC_2). Overall, the interrater reliability was very good. For RSVP Past, Recent, and Relevance items, the average interrater reliability for ICC_1 was .81 (range = .43 to .95), .91 (range = .71 to .98) and .83 (range = .49 to .95) respectively. For RSVP Past, Recent, and Relevance items, the average interrater reliability for ICC_2 was .90 (range = .60 to .97), .96 (range = .83 to .99) and .91 (range = .66 to .97) respectively. The interrater reliability for summary risk ratings was excellent (ICC_1 range = .95 to 1.00, ICC_2 range = .98 to 1.00).

RSVP Predictive Validity

Very little is known about the predictive validity of the RSVP. Based on a sample of 53 sex offenders, Kropp (2001) looked at the relationship between RSVP ratings (total scores and Case Prioritization ratings) and recidivism. Although RSVP total scores were not significantly correlated with sexually violent recidivism (r = .23), Case Prioritization ratings were significantly correlated with sexually violent recidivism (r = .40, p < .05). When compared to actuarial risk measures, the RSVP was more highly correlated with future sexual violence. For example, Static-99 total scores were correlated with sexual recidivism, r = .30, p < .05.

Darjee et al. (2016) examined the predictive validity and utility of the RSVP in their sample of 109 people. The follow-up varied between six months and five years, with an average of 3.25 years. With regards to predictive validity, the results were mixed. The RSVP Case Prioritization predicted sexual recidivism, according to survival analysis. Furthermore, total scores and summary risk ratings predicted serious and violent offending, whether sexual or non-sexual. However, using ROC analyses, RSVP total scores and summary risk ratings, including Case Prioritization, did not predict sexual recidivism. This study also found that the level of intervention moderated the association between risk and recidivism: those who did not receive high levels of intervention or management reoffended more quickly and more often.

RSVP Utility

An early review of the RSVP was done by Prescott (2004), who evaluated the RSVP in terms of its usefulness and appropriateness as a sexual violence risk assessment tool. In this review, the RSVP was described as providing a useful framework for sexual violence risk assessment that goes beyond labelling an offender's level of risk for reoffending. According to Prescott, one of the main benefits of the RSVP is that it identifies risk factors and asks the assessor to determine the relevance of each risk factor to future sexual offending, which allows for a more complete risk-formulation and assessment-based treatment plan. With regards to limitations, Prescott suggested that the RSVP could be improved by the inclusion of additional information, such as individualized treatment needs based on the type of offender and a framework for conceptualizing and communicating an offender's risk over a lifetime. Other suggestions included having the RSVP provide current perspectives in the literature on acute versus stable dynamic risk factors, as well as recommendations for integrating the varied perspectives in risk assessment. Finally, the RSVP was described as lacking a discussion on anchoring risk assessments using base rate information and with yet-to-be-determined interrater reliability and predictive validity.

A number of recent studies have empirically evaluated the usefulness of the RSVP. For example, Judge, Quayle, O'Rourke, Russell, and Darjee (2014) explored the qualitative experience of RSVP users. The participants were 31 criminal justice professionals in Scotland. Semi-structured interviews were conducted with the

participants, who also completed questionnaires. The results indicated that the participants valued the RSVP assessment process as reflected in five themes that emerged from the data: informing risk management recommendations; confirming what was known and giving weight to risk factors; understanding the implications of personality; targeting treatment needs; and the usefulness and limitations of risk assessment. With regards to treatment, participants found the assessments particularly helpful in accessing services that may not have otherwise been available for the sexual offender. This study lends support for the RSVP having added value over actuarial risk assessment measures in terms of looking beyond "predicting" recidivism to understanding how to best manage and treat sexual offenders. As noted earlier, the basis of management and treatment recommendations should be case formulation, which should provide a coherent picture of how the convergence of offenders' personality characteristics and risk factors place them at risk for sexually offending. Although the area of case formulation has only begun to receive attention, in the study discussed earlier by Wilson (2013), case formulation based on the RSVP was examined. The findings of this study suggested that there was modest consistency between the case formulations for the same cases across different evaluators.

Darjee et al. (2016) examined the concurrent validity and utility of the RSVP in their sample of 109 people. The RSVP had good concurrent validity with respect to other violence risk assessment measures, including the Risk Matrix 2000 (RM2000; Thornton, 2010) and Hare Psychopathy Checklist-Revised (PCL-R; Hare, 1991, 2003). Also, the RSVP risk scenarios developed by assessors were found to closely resemble the actual recidivism events with respect to victim type and severity of harm.

Current Study

Insofar as the RSVP may be considered a parallel form of the SVR-20, its predictive validity is assumed to be similar to that of the SVR-20. However, to date there are only two studies (Darjee et al., 2016; Kropp, 2001) looking at the predictive value of RSVP ratings. Given the paucity of research evaluating the reliability and validity of the RSVP, the purpose of this research was to examine the psychometric properties of the RSVP, vis-à-vis the SVR-20 and a number of actuarial sexual violence risk assessment

instruments. The sample consisted of 100 sexual offenders who had completed a community-based sexual offender treatment program in British Columbia, Canada.

The research questions guiding this project were:

a. What is the interrater reliability of risk judgements made using the RSVP? Specifically, what is the interrater reliability of (a) presence and relevance ratings for RSVP risk factors (individual, domain, and total), and (b) summary judgements?
b. What is the concurrent validity of the RSVP vis-à-vis the SVR-20 and actuarial risk assessment instruments, including the Static-99R, Static-2002R, and SORAG? Specifically, what is the concurrent validity of RSVP (a) risk factors (individual, domain, and total), and (b) summary judgements with respect to corresponding ratings on the SVR-20 and actuarial risk assessment instruments?
c. What is the predictive validity of the RSVP vis-à-vis sexual recidivism? Specifically, what is the predictive validity of RSVP (a) risk factors (domain, total), and (b) summary judgements with respect to sexual recidivism?
d. What is the predictive validity of the RSVP relative to the SVR-20 and actuarial risk assessment instruments? Specifically, what is the predictive validity of RSVP (a) risk factors (domain, total), and (b) summary judgements with respect to corresponding ratings on the SVR-20 and actuarial risk assessment instruments?

METHOD

Participants

Participants[8] were 100 male sex offenders who had engaged in an outpatient sex offender treatment program between 2002 and 2003. The treatment program was provided by the Forensic Psychiatric Services Commission (FPSC) in British Columbia, Canada. All of the participants were mandated to participate in the FPSC treatment program as a part of their conditions of community supervision. The data were collected via an archival file review and therefore, informed consent was not obtained from the participants. To ensure the protection of participants' personal information, a rigorous and thorough ethics approval process was undertaken including Simon Fraser University, the Forensic Psychiatric Services Commission of British Columbia, and the Vancouver Police Department. Of the original 116 files that were reviewed, 15 files were excluded due to insufficient file information (e.g., incomplete or missing psychosocial or criminal history) and one file was excluded due to the participant being female, leaving a sample of 100 participants.

Sample Characteristics

The mean age of participants was 41 years (*SD* = 12.41 years, range = 19 to 77) at the completion of the FPSC treatment program. In terms of ethnicity, 66% of the participants were Caucasian, 15% were Aboriginal, and 19% were of other ethnicities.

[8] The term participants is used with the acknowledgment that the individuals comprising the sample did not actually "participate" in the study.

Both marital and employment status were considered at the time of the index offence (the sexual offence that led to participants' being referred to the FPSC treatment program). Approximately half (48%) of the participants were in a live-in relationship (married or common-law), with the remaining 52% being either single, separated, divorced, or widowed. The majority of the sample (61%) was employed either full or part-time. With respect to psychosocial history, 60% of participants had a history of serious alcohol or drug abuse, and 25% had a diagnosis of a major mental disorder.

Participants were categorized based on their sexual offending history. This sample included 21% incest offenders (intra-familial victims only), 26% pedophiles (extra-familial victims under age 16), 23% adult only victim offenders (intra- or extra-familial victims age 16 or older), and 30% mixed offenders (victims both under age 16 and age 16 or older, either intra- or extra-familial). The index offence was the first sexual offence conviction for 61% of participants and the first conviction for any type of offence for 36% of participants. In other words, 39% of participants had been convicted of a sexually violent offence in the past and 64% had a criminal conviction of some kind (including sexual violence) in the past.

The FPSC treatment program was 13 weeks in length and consisted of weekly, three-hour group meetings that were psychoeducational in nature. Some participants also engaged in individual sessions over the course of treatment. The mean length of treatment was 3.04 months (*SD* = 1.27, range = .95 to 7.16). The length of treatment varied because of a number of factors such as whether the participant engaged in individual treatment or if the participant took longer to complete the treatment program (i.e., multiple attempts). Prior to the FPSC treatment program, 45% of participants had already engaged in some form of sex offender treatment program. This percentage is higher than the proportion (39%) of offenders who had a previous sexual offence conviction, due to a number of first-time sexual offenders having participated in sex offender treatment while incarcerated for their current offence, prior to participating in the FPSC treatment program.

Procedure

The study utilized a retrospective follow-up or quasi-prospective design. Various risk measures were coded retrospectively from Forensic Psychiatric Services Commission files in about 2005 to 2006. This was after participants had completed the FPSC treatment program (between late 2002 and early 2004), but well before the end of the 10 year follow-up period (in late 2012) and the subsequent collection and coding of outcome data (in 2013 and 2014). All of the risk measures were coded based on a thorough review of participants' files as they existed at the end of the FPSC treatment program, blind to any file information that may have been added after completion of treatment. The file information included criminal records, police reports for index and past offences, pre-sentence reports, psychological and risk assessment reports, notes from previous treatment, and FPSC treatment program summaries.

Each participant's file was coded by three members from a team of four researchers. All four researchers were senior graduate students in forensic psychology or criminology who had extensive training or experience with the respective risk measures. For each participant, one researcher coded the PCL:SV and all the risk measures (RSVP, SVR-20, Static-99/99R, Static-2002/2002-R, and SORAG); a second researcher coded only the PCL:SV and the structured professional judgement risk measures (RSVP and SVR-20); and a third researcher coded only the actuarial risk assessment instruments (Static-99/99R, Static-2002/2002-R, and SORAG). All researchers coded files blind to the ratings made by other researchers, and also blind to (indeed, several years before) outcome. Thus, I was able to evaluate the interrater reliability of decisions made using the PCL:SV and all the risk measures under conditions where one of the researchers made ratings using structured professional judgement guidelines blind to ratings on the actuarial risk assessment measures. Similarly, another researcher made ratings on the actuarial risk assessment measures blind to ratings made using structured professional judgement guidelines.

After coding the files independently, researchers then met in pairs to discuss their findings and made a final set of consensus ratings that were used in concurrent and predictive validity analyses. These consensus ratings were intended to minimize the

impact of interrater unreliability—and, consequently, maximize statistical power—in the validity analyses.

The recidivism data were coded from multiple electronic police and criminal record databases. The recidivism data were collected by a sworn member of the Vancouver Police Department, who was blind to the content of the reviewed FPSC files and the ratings on the risk measures. The beginning of the follow-up period for the recidivism data was the participants' completion in the FPSC treatment program, between September 2002 and March 2004. The follow-up period ended in December 2012, resulting in a follow-up period of up to 10 years.

Six participants died during the follow-up period. The date of death was known for four of the participants, and this date was used as the end of their follow-up. The date of death was unknown for the other two participants, but for both I was able to determine dated of last known contact with law enforcement, corrections, or FPSC staff, which was used as the end of their follow-up.

Materials

Risk for Sexual Violence Protocol (RSVP)

The RSVP (Hart et al., 2003) is a set of structured professional judgement guidelines for completing sexual violence risk assessments. The RSVP comprises 22 individual risk factors from the following five domains: History of sexual violence, Psychological adjustment, Mental disorder, Social adjustment, and Manageability (see Table 1). For a complete overview of the RSVP coding sheets, please see Appendix A. Presence ratings were made for each individual risk factor for two time periods: past, that is, more than one year prior to the completion of the FPSC treatment program; and recent, that is, within the year prior to the completion of the FPSC treatment program. Presence ratings are coded on a 3-point ordinal scale (*Absent*, *Possibly/partially relevant*, *Present*). To reduce the number of analyses and facilitate direct comparison with the other risk measures, I combined the "past" and "recent" ratings to create a single Presence (i.e., "ever present") rating by taking the highest of the past or recent ratings for each individual risk factor.

Relevance ratings were also made for the 22 individual risk factors. These ratings reflected the researcher's judgements of the functional relevance of each risk factor with respect to potential perpetration of sexual violence in a given case. Relevance ratings are also made on a 3-point ordinal scale (*Not relevant*, *Possibly/partially relevant*, *Relevant*).

Finally, Summary Risk ratings were made for each participant. These are integrative judgements about the risks posed by each individual in light of the presence and relevance of the individual risk factors. I evaluated three of the Summary Risk ratings: Case Prioritization, reflecting the overall level of effort or intervention required to prevent future sexual violence; Serious Harm, reflecting the likelihood that any future sexual violence perpetrated by the participant might result in serious or life-threatening physical injury; and Immediate Action, reflecting the likelihood that any future sexual violence perpetrated by the participant might occur in the near future. Summary Risk ratings were made on a 3-point ordinal scale (*Low/routine/no*, *Moderate/elevated/possible*, *High/urgent/yes*).

For the purpose of data analyses, the RSVP ratings were numerically re-coded: for Presence, 0 = *Absent*, 1 = *Possibly/partially present*, 2 = *Present*; for Relevance, 0 = *Not relevant*, 1 = *Possibly/partially relevant*, 2 = *Relevant*; and for Summary Risk, 0 = *Low/routine/no*, 1 = *Moderate/elevated/possible*, 2 = *High/urgent/yes*. Recoding RSVP ratings in this manner is consistent with previous research involving the RSVP (e.g., Hart, 2003b; Hart et al., 2008; Watt et al., 2006; Watt & Jackson, 2008) and other structured professional judgement risk assessment instruments such as the SVR-20 (e.g., Barbaree et al., 2008; de Vogel et al., 2004; Stadtland et al., 2005). The numerical ratings were summed to create composites, including total scores (sum of all risk factors) and domain scores (sum of risk factors within each domain).

As the interrater reliability and distribution of the RSVP ratings was a focus of the current research, the relevant findings are presented in the Results section.

Table 1 RSVP Risk Factors

Domain	Risk Factor
History of sexual violence	1. Chronicity of sexual violence
	2. Diversity of sexual violence
	3. Escalation of sexual violence
	4. Physical coercion in sexual violence
	5. Psychological coercion in sexual violence
Psychological adjustment	6. Extreme minimization or denial of sexual violence
	7. Attitudes that support or condone sexual violence
	8. Problems with self-awareness
	9. Problems with stress or coping
	10. Problems resulting from child abuse
Mental disorder	11. Sexual deviance
	12. Psychopathic personality disorder
	13. Major mental illness
	14. Problems with substance use
	15. Violent or suicidal ideation
Social adjustment	16. Problems with intimate relationships
	17. Problems with non-intimate relationships
	18. Problems with employment
	19. Nonsexual criminality
Manageability	20. Problems with planning
	21. Problems with treatment
	22. Problems with supervision

Sexual Violence Risk-20 (SVR-20)

The SVR-20 (Boer et al., 1997) is a set of structured professional judgement guidelines for conducting sexual violence risk assessments. The SVR-20 is the precursor to the RSVP, and therefore, is very similar to the RSVP in terms of the risk factors that are considered (the RSVP includes 18 risk factors that directly parallel those in the SVR-20). However, the RSVP added some risk factors that reflect aspects of psychological adjustment and also has a more complex administration procedure intended to facilitate the development of management and treatment plans.

The SVR-20 comprises 20 individual risk factors from four domains: Psychological adjustment, Social adjustment, History of sexual offences, and Manageability (see Table 2 for the risk factors and Appendix B for the SVR-20 coding sheet). Presence ratings ("ever present") for each risk factor were made using a 3-point ordinal scale (*Absent*, *Possibly/partially*, *Present*). The ordinal ratings were re-coded numerically, 0 = *Absent*, 1 = *Possibly/partially present*, 2 = *Present*. The individual risk factor ratings were summed to create composites, including a total score and domain scores.[9]

Overall, the interrater reliability of consensus SVR-20 ratings was excellent. For the risk measures other than the RSVP, I focus on the interrater reliability of the consensus ratings that were actually used in the validity analyses, and the best estimate of the effective interrater reliability of the consensus ratings is ICC_2. Details concerning the interrater reliabilities of individual risk factors are provided in the Appendix (see Table G1). Briefly, they ranged from ICC_2 = .78 to .98, with *M* = .89 and *Mdn* = .89. For the total and domain scores, the interrater reliabilities were as follows: total, ICC_2 = .96; Psychological adjustment, ICC_2 = .96; Social adjustment, ICC_2 = .94; History of sexual offences, ICC_2 = .94; and Future plans, ICC_2 = .86.

[9] The SVR-20 also permits evaluators to make Summary Risk ratings to reflect integrative global judgements of the risks posed by participants; but as I expected they would be substantively identical to those made using the RSVP, I did include these ratings in the research.

Table 2 SVR-20 Risk Factors

Domain	Risk Factor
Psychological adjustment	1. Sexual deviation
	2. Victim of child abuse
	3. Psychopathy
	4. Major mental illness
	5. Substance use problems
	6. Suicidal/homicidal ideation
Social adjustment	7. Relationship problems
	8. Employment problems
	9. Past nonsexual violent offences
	10. Past nonviolent offences
	11. Past supervision failure
History of sexual offences	12. High density
	13. Multiple types
	14. Physical harm
	15. Weapons/threats
	16. Escalation in frequency or severity
	17. Extreme minimization/denial
	18. Attitudes that support or condone
Future plans	19. Lacks realistic plans
	20. Negative attitude toward intervention

For descriptive purposes, the distribution of consensus ratings for the SVR-20 individual risk factors are presented in Appendix G, Table G1. Overall, 14 out of 20 were rated as present to some degree (scores of at least 1) in 50% or more of cases. The distribution of composite (total and domain) scores for the SVR-20 are summarized in Table 3. The correlation between the total and domain score is presented in Table 4.

Table 3 Distribution of SVR-20 Domain and Total Scores

Scores	*M*	*(SD)*	Min	Max
History of sexual offences	5.50	(2.64)	0	12
Psychological adjustment	5.34	(2.66)	0	12
Social adjustment	6.00	(2.73)	1	10
Future plans	2.79	(1.19)	0	4
Total	19.63	(6.92)	5	35

Note. *N* = 100.

Table 4 Correlation (r) Between SVR-20 Total and Domain Scores

	Total	SEX	PSY	SOC
SEX	.69***	1.0		
PSY	.81***	.38***	1.0	
SOC	.80***	.25*	.58***	1.0
FUT	.63***	.35***	.31**	.52***

Note. *N* = 100. SEX = History of sexual offences, PSY = Psychological adjustment, SOC = Social adjustment, FUT = Future plans.

Static-99/Static-99R

The Static-99 (Hanson & Thornton, 1999) is an actuarial risk assessment instrument used for estimating risk for sexual and violent recidivism among adult male sexual offenders. As noted in the Introduction, the Static-99 was revised in 2011 and the revised version is recommended for sexual violence risk assessment over the Static-99. The only change reflected in the Static-99R is a more accurate account of the relationship between an offender's age and risk for future sexual violence. As with the Static-99, the Static-99R features 10 items of a static or historical nature that have empirical support for their relationship to future sexual and non-sexual violence. The 10 items are outlined in Table 5 and the Static-99R coding sheet can be found in Appendix

C. As indicated in Table 5, the Static-99R items are categorized into five domains: Sexual deviance, Range of potential victims, Persistence, Antisocial, and Age.

The 10 items' scores are weighted and summed to produce a total score that can range from -3 to 12. Total scores correspond with four suggested risk categories (*Low* = -3 to 1, *Low-moderate* = 2 or 3, *Moderate-high* = 4 or 5, *High* = $\geq$6). Total scores are associated with percentile ranges, relative risk ratios and sexual recidivism rates. For example, according to the Static-99R Evaluators' Manual (Phenix et al., 2015) a total score of 3 would place an individual in the *Low-moderate* risk category (between the 57th and 74th percentile) relative to other adult male sexual offenders. An offender with this score would be estimated to have a recidivism rate 1.4 times higher than the recidivism rate of a typical sexual offender. The Static-99R normative data includes two different samples, routine and high risk/needs offenders. On average, routine sex offenders with a total score of 3 sexually recidivate at a rate of 8% over 5 years. High risk/needs sex offenders with this total score, on average, sexually recidivate at a rate of 14% over 5 years and 23% over 10 years.

Table 5 Static-99R Risk Factors

Domain	Risk Factor
Sexual deviance	1. Male victim(s)
	2. Never lived with a romantic partner for at least 2 years
	3. Non-contact sexual offence (conviction)
Range of potential victims	4. Unrelated victim(s)
	5. Stranger victim(s)
Persistence	6. Prior sexual offences (charges and/or convictions)
Antisocial	7. Current non-sexual violence
	8. Prior conviction(s) for non-sexual violence
	9. 4+ sentencing dates
Age	10. Under age 35

For this study, the Static-99 was originally used and coded based on the revised coding rules (Harris et al., 2003). At the time of the original coding, the Static-99R was not available. Given that the Static-99R is recommended over the Static-99, the Static-99 scores were recoded into Static-99R scores, which were used in the analyses.

The interrater reliability of the individual Static-99R items is presented in Appendix G, Table G3[10]. The ICC_2 values for items ranged from .67 to .96 (*M* = .87, *Mdn* = .89). The interrater reliability of the consensus Static-99R total scores and risk categories was excellent, ICC_2 = .96 and .94, respectively (please see Appendix G, Table G3).

The distribution of Static-99R item scores is presented in Appendix G, Table G2. Overall, 1 of the 10 items was rated as present (score of at least 1) for 50% or more of the participants: Any unrelated victims. The distribution of Static-99R total scores ranged

[10] Once again, I will focus on consensus ratings here, which were those analysed in the Results section; the effective interrater reliability of the consensus ratings was estimated using ICC_2.

from -3 to 10, *M* = 2.76, *SD* = 2.82. With respect to risk categories, 28% of participants fell in the *Low* category, 33% in *Low-moderate*, 24% in *Moderate-high*, and 15% in *High*. The correlation between total scores and risk categories was very high, *r* = 95, *p* < .001.

Static-2002/Static-2002R

The Static-2002 (Hanson & Thornton, 2003) is an actuarial risk assessment instrument developed for the purpose of estimating relative risk for violent (sexual and nonsexual) and nonviolent recidivism among adult male sexual offenders. The Static-2002 was developed from the Static-99, thus explaining the similarity between these two measures (the Static-99 overlaps with nine items on the Static-2002). The major coding differences between the Static-99 and Static-2002 was the deletion of two items and the addition of six new items, resulting in a total of 14 items, as well as the addition of a *Moderate* risk category. The Static-2002 was revised (Helmus et al., 2012) at the same time as the Static-99, and consistent with the Static-99R, it is recommended that the Static-2002R is used over the Static-2002 for the purpose of risk assessment. The only difference is that the Static-2002R codes age in greater detail. The Static-2002R items are grouped into five domains: Age, Persistence of sex offending, Deviant sexual interests, Relationship to victims, and General criminality. See Table 6 for the individual items and Appendix D for the Static-2002R coding sheet.

The individual items on the Static-2002R are coded, weighted, and summed to produce a total score that can range from -2 to 13. Total scores are then divided into five risk categories: *Low* = -2 to 2, *Low-moderate* = 3 or 4, *Moderate* = 5 or 6, *Moderate-high* = 7 or 8, and *High* = $\geq$9. Each risk category also has estimated percentiles, relative risk ratios, and sexual recidivism rates. For example, according to the Static-2002R Evaluators' Manual (Phenix et al., 2015), a total score of 5 would place an individual in the *Moderate* risk category (between the 71st and 85th percentile), relative to other adult male sexual offenders. An offender with this score would be estimated to have a recidivism rate 1.9 times higher than the recidivism rate of a typical sexual offender. The Static-2002R normative data includes two different samples, routine and high risk/needs offenders. On average, routine sex offenders with a total score of 5 sexually recidivate at a rate of 14% over 5 years. High risk/needs sex offenders with this total score, on average, sexually recidivate at a rate of 19% over 5 years.

Table 6 Static-2002R Risk Factors

Domain	Risk Factor
Age	1. Age at release (< 40 years)
Persistence of sexual offending	2. Prior sentencing occasion(s) for sexual offences
	3. Juvenile arrest for a sexual offence & adult conviction for a separate sexual offence
	4. Rate of sexual offending
Deviant sexual interests	5. Previous sentence for a non-contact sexual offence
	6. Any male victim(s)
	7. Two or more victims <12, one unrelated to offender
Relationship to victim	8. Unrelated victim(s)
	9. Stranger victim(s)
General criminality	10. Any prior involvement with criminal justice system
	11. Prior sentencing occasion for anything
	12. Community supervision violation(s)
	13. Years free prior to index sexual offence
	14. Previously sentenced for non-sexual violence

For this study, the Static-2002 was originally used as the Static-2002R was not available. Given that the Static-2002R is recommended over the Static-2002, the scores were recoded into Static-2002R scores, which were used in the analyses.

The interrater reliability of Static-2002 item scores is presented in Appendix G, Table G4.[11] ICC_2 values ranged from .71 to .98 (*M* = .86, *Mdn* = .85). The interrater reliability of the Static-2002R total scores and risk categories was excellent, ICC_2 = .96 and .93, respectively (please see Appendix G, Table G2).

[11] Once again, I will focus on consensus ratings here, which were those analysed in the Results section; the interrater reliability of the consensus ratings was indexed using ICC_2.

The distribution of Static-2002R item scores is presented in Appendix G, Table G4. Overall, 2 out of 14 items were present (score of at least 1) for 50% or more of the participants: Any unrelated victims and Any prior involvement in the criminal justice system. The distribution of Static-2002R total scores ranged from -2 to 11, *M* = 4.32, *SD* = 2.91. With respect to risk categories, 25% fell in the *Low* category; 25% in *Low-moderate*; 29% in *Moderate*; 12% in *Moderate-high*; and 9% in *High*. The correlation between total scores and risk categories was very high, $r = 95$, $p < .001$.

Sex Offender Risk Appraisal Guide (SORAG)

The SORAG (Quinsey, Harris, Rice, & Cormier, 1998, 2006; Harris, Rice, Quinsey & Cormier, 2015) is an actuarial risk assessment instrument for evaluating the risk for sexual violence among adult male offenders who have committed at least one previous contact sexual offence. The SORAG comprises 14 items that are coded, weighted, and summed to produce a total score that can range from -26 to +51. Missing items on the SORAG are given a score of 0. For this study, one of the items (the phallometric test results) was omitted for all participants due to missing data. See Table 7 for individual items and Appendix E for the SORAG coding sheets. SORAG total scores correspond with nine bins that are ranked 1 to 9, with Bin 9 representing the highest risk (no descriptive labels are provided). Risk categories are associated with percentile rankings and estimated probabilities for violent recidivism, relative to the sample of sexual offenders used in the development of the SORAG. For example, according to the most recent norms available (Harris et al., 2015), an offender with a total SORAG score of 20 would fall into Bin 7, which corresponds with the 70th percentile and recidivism rates of 51% over 5 years and 76% over 12 years.

Table 7 SORAG Risk Factors

	Risk Factor
1.	Lived with both biological parents to age 16 (except for death of parent)
2.	Elementary school maladjustment
3.	History of alcohol problems
4.	Marital status at time of index offence
5.	Charges and convictions for non-violent offences prior to index offence
6.	Charges and convictions for violent offences prior to index offence
7.	Number of convictions for sexual offences prior to index offence
8.	History of sex offences against females under age 14 only
9.	Failure of prior conditional release
10.	Age at index offence
11.	Meets diagnostic criteria for any personality disorder
12.	Meets diagnostic criteria for schizophrenia
13.	Psychopathic personality disorder

The interrater reliability of SORAG item scores is presented in Appendix G, Table G5.[12] ICC_2 values ranged from .66 to .96 (*M* = .83, *Mdn* = .82). The interrater reliability of the SORAG total scores and bins was excellent, both ICC_2 = .96 (please see Appendix G, Table G2).

The distribution of SORAG item scores is also presented in Appendix G, Table G4. Overall, 3 out of 14 items were present (score of at least 1) for 50% or more of the participants: Does not meet criteria for schizophrenia, History of female victims under age 14, and Did not live with biological parents. SORAG total scores ranged from -19 to 35, *M* = 4.69, *SD* = 13.60. With respect to risk categories, the distribution was as follows:

[12] Once again, I will focus on consensus ratings here, which were those analysed in the Results section; the interrater reliability of the consensus ratings was indexed using ICC_2.

Bin 1 = 16%, Bin 2 = 18%, Bin 3 = 13%; Bin 4 = 12%; Bin 5 = 20%; Bin 6 = 4%; Bin 7 = 8%; Bin 8 = 5%; and Bin 9 = 4%. The correlation between SORAG total scores and bins was very high, $r = .99$, $p < .001$.

The Psychopathy Checklist: Screening Version (PCL:SV)

Several of the risk measures used in this study contain a risk factor or item that reflects psychopathic personality disorder, in particular as measured by the PCL-R or PCL:SV. I used the PCL:SV to measure overall severity of symptoms of psychopathy because it is particularly appropriate when making file-based ratings and can be used to estimate corresponding PCL-R scores (Cooke, Michie, Hart, Hare, 1999; Hart et al., 1995).

The PCL:SV comprises 12 items (see Table 8). The items are coded on a 3-point numerical scale (0 = *Absent*, 1 = *Some indication*, 2 = *Present*). Item scores are unit-weighted and summed to yield a total score, which can range from 0 to 24. Total scores can be divided into three categories to reflect overall severity of symptoms of psychopathy: 0 to 12 = *No/low*, 13 to17 = *Possible/moderate*, and $\geq$ 18 = *Definite/high* (Hart et al., 1995).

The interrater reliability of the consensus PCL:SV total scores was excellent, $ICC_2 = .95$.

Consensus total scores ranged from 2 to 24, $M = 10.72$, $SD = 5.52$. With respect to categorical ratings, 63% of the participants fell into the *No/low* group, 14% fell into the *Possible/moderate* group, and 13% fell into the *Definite/high* group.

Table 6 PCL:SV Items

Risk Factor
1. Superficial
2. Grandiose
3. Deceitful
4. Lacks Remorse
5. Lacks Empathy
6. Does Not Accept Responsibility
7. Impulsive
8. Poor Behavioral Controls
9. Lacks Goals
10. Irresponsible
11. Adolescent Antisocial Behavior
12. Adult Antisocial Behavior

Recidivism Data

Recidivism data were collected through the Vancouver Police Department and involved a comprehensive search including the following federal, provincial and municipal databases: Canadian Police Information Centre (CPIC), Criminal Number Index (CNI), Police Records Information Management Environment (PRIME), Police Information Records System (PIRS), Law Enforcement Information Portal (allows access to PRIME databases across Canada in those jurisdictions that use PRIME), and the Motor Vehicle Branch (for driver's licenses, date of death for deceased participants, etc.). In previous recidivism research (e.g., Barbaree, 2005), the use of CPIC records was viewed as a methodological improvement over correctional or parole board file information. However, the information in CPIC is not up-to-date. Therefore, the inclusion of provincial and municipal police record databases (which are more current) allowed for more recent and complete recidivism data.

I coded recidivism according to three definitions: *Any new conviction*, *Any new charge or conviction*, and *Any new police contact* for a criminal offence during the follow-

up period. I further subdivided each of these types of recidivism as involving *Sexual violence*, *Nonsexual violence*, or *Nonviolent criminality*, according to the information in the databases reviewed. The rates of these six different types of recidivism are presented in Table 9. As the table indicates, the rates for *Any new police contact* were about twice as high as those for *Any new charge or conviction*, which in turn were about twice as high as those for *Any new conviction*. As the table also indicates, the rate for *Nonviolent criminality* was about twice that of the rates for *Sexual violence* and *Nonsexual violence*.

Table 9 Rates of Recidivism

	Any New Conviction	Any New Charge or Conviction	Any New Police Contact
Sexual violence	5%	12%	29%
Nonsexual violence	6%	17%	32%
Nonviolent criminality	18%	32%	62%

Note. $N = 100$

Although most studies tend to define recidivism in terms of new convictions (e.g., Doren, 2002; Hanson & Bussière, 1998) or new charges and convictions (e.g., Harris et al., 2015; Olver & Wong, 2006; Smid et al., 2014), I decided to use the following definition of recidivism: *Any new police contact involving sexual violence*. This definition includes anyone who was charged with or convicted of a new offence involving sexual violence. This definition was utilized so as to minimize potential false negative errors as it includes those who are being investigated for sexual violence and are not yet charged or convicted. It also includes those whose official charge or conviction was for a nonsexual or nonviolent offence despite evidence of sexually motivated violence. Therefore, this broader definition minimizes the problem that sexual violence tends to be under-reported and may have a low rate of clearance via charge or conviction (e.g., Lievore, 2003; Prentky, Lee, Knight, & Cerce, 1997). But this definition almost certainly includes some false positive errors, that is, cases in which participants were investigated but in fact did not perpetrate new sexual violence. Overall, I believe the definition I selected reasonably balances the chances of false negative versus false positive errors.

To calculate maximum time at risk, I subtracted the date on which each participant completed the FPSC treatment program from the date of the criminal record check (or the date of death, whichever came first). The average maximum time at risk was almost 10 years (*M* = 9.50, *SD* = .42, range = 8.69 to 10.08 years). To calculate time to failure, I subtracted the date on which each participant completed the FPSC treatment program from the date of the first occurrence of *Any new police contact involving sexual violence* (or the date of death, whichever came first).

Data Analyses

The data analyses for this research was conducted using the Statistical Package for the Social Sciences, Version 18, and Stata SE, Version 13.1.

Descriptive Results

For each of the sexual risk assessment instruments, the frequency of the individual risk factors was included. Pearson correlations (*r*) were calculated between the total score and risk categories for each risk assessment instrument.

Interrater Reliability

The interrater reliability of all of the sexual violence risk assessment instruments was indexed using ICC_1, calculated for absolute agreement using a mixed effects model. In addition, the estimated reliability of the consensus ratings was indexed using ICC_2. As noted previously, given that most analyses were based on consensus ratings, ICC_2 should be considered the most relevant index of IRR in this study.

Concurrent validity

Concurrent validity was evaluated by calculating Pearson correlations (*r*) between the RSVP and the other sexual violence risk assessment measures.

Predictive validity

The predictive validity of the RSVP and the other sexual violence risk assessment instruments was indexed using a number of different types of analyses. The predictive validity of risk assessment instruments was indexed using AUC (Hanley & McNeil, 1982). Within the present sample, an AUC value represents the probability that a randomly selected sexual recidivist will have a higher score on a particular risk assessment variable than a randomly selected sexual nonrecidivist. AUC values fall between 0 and 1. An AUC value of .5 represents prediction at the level of chance and an AUC value of 1 represents perfect prediction, with AUC values of .64 to .70 having been defined as being equivalent to a medium effect size and AUC values $\geq$ .71 being associated with large effects (Rice & Harris, 2005).

Next, predictive validity was evaluated using survival analysis, which has two major advantages relative to AUC analysis: first, it takes into account time to failure; and second, it takes into account censored cases. I examined the predictive validity of Case Prioritization ratings using Kaplan-Meier survival analysis, which is a non-parametric method appropriate for evaluating categorical independent variables. Kaplan-Meier survival analysis is often used in recidivism research (e.g., Lund, Hofvander, Forsman, Anckarsäter, & Nilsson, 2013; Stadtland et al., 2013). For the remaining analyses, I used Cox regression survival analysis, which is a semi-parametric method appropriate for evaluating both categorical and continuous independent variables. Cox regression survival analysis has been used in previous recidivism research (e.g., Fazel, Sjöstedt, Långström, Grann, 2006; Hanson, 2005; Harris et al., 2003; Olver & Wong, 2006).

RESULTS

Distribution of RSVP Ratings

The distribution of RSVP Presence and Relevance ratings for individual risk factors is presented in Table 10. Looking first at Presence ratings, 12 out of 22 risk factors were coded as present (score of 2) for 50% or more of the participants, and only 1 risk factor was coded as present for more than 85% of the participants. Turning next to Relevance ratings, 8 out of 22 risk factors were coded as relevant (score of 2) for 50% or more of the participants, and none had an extreme distribution. In general, the Presence ratings for a given risk factor tended to be slightly more extreme than the corresponding Relevance ratings. For both Presence and Relevance ratings, the five risk factors most frequently endorsed (score of 2 in $\geq$ 70% of cases) were: Problems with self-awareness; Problems with stress and coping; Problems with intimate relationships; Problems with non-intimate relationships; and Problems with Planning.

The distribution of RSVP Presence and Relevance total and domain scores is summarized in Table 11. The scores all were distributed quite normally, with a *M* around the centre of the theoretical range and substantial variability. Once again, the corresponding scores are similar but slightly higher for Presence than for Relevance.

The distribution of RSVP Summary Risk ratings is summarized in Table 12. Few participants received a high score (i.e., 2) on these ratings, which is generally consistent with the distribution of scores on the other risk measures and indicates that the sample maybe characterized overall as low or low-moderate risk in terms of the likelihood, severity, and imminence of potential future sexual violence.

Table 10 Distribution of RSVP Presence and Relevance Ratings, Individual Risk Factors

	Presence			Relevance		
RSVP Risk Factors	**0**	**1**	**2**	**0**	**1**	**2**
Chronicity of sexual violence	26	12	62	30	16	54
Diversity of sexual violence	64	25	11	67	15	18
Escalation of sexual violence	54	23	23	62	19	19
Physical coercion in sexual violence	58	13	29	59	16	25
Psychological coercion in sexual violence	30	23	47	30	24	46
Extreme minimization or denial of sexual violence	10	25	65	15	20	65
Attitudes that support or condone sexual violence	15	35	50	19	35	46
Problems with self-awareness	1	11	88	1	17	82
Problems with stress or coping	5	22	73	6	23	71
Problems resulting from child abuse	37	11	52	44	25	31
Sexual deviance	40	28	32	41	26	33
Psychopathic personality disorder	63	24	13	63	24	13
Major mental illness	45	30	25	48	33	19
Problems with substance use	28	12	60	37	17	46
Violent or suicidal ideation	45	24	31	60	22	18
Problems with intimate relationships	3	17	80	3	14	83
Problems with non-intimate relationships	4	14	82	8	16	76
Problems with employment	18	35	47	28	27	45
Nonsexual criminality	35	33	32	55	27	18
Problems with planning	3	24	73	5	23	72
Problems with treatment	14	25	61	15	25	60
Problems with supervision	41	9	50	48	17	35

Note. N = 100. 0 = *Not Present or Relevant*; 1 = *Possibly/partially Present or Relevant*; 2 = *Present or Relevant.*

Table 11 ***Distribution of RSVP Presence and Relevance Ratings, Domain and Total Scores***

	Presence				Relevance			
Scores	***M***	***(SD)***	**Min**	**Max**	***M***	***(SD)***	**Min**	**Max**
History of sexual violence	4.40	(2.55)	0	10	4.14	(2.65)	0	10
Psychological adjustment	7.60	(1.78)	2	10	7.10	(1.87)	2	10
Social adjustment	5.81	(1.87)	1	8	5.28	(1.82)	1	8
Manageability	4.26	(1.83)	0	6	3.99	(1.86)	0	6
Mental disorder	4.40	(2.22)	0	10	3.80	(2.08)	0	9
Total	26.47	(7.23)	10	42	24.31	(7.52)	9	42

Note. $N = 100$.

Table 12 ***Distribution of RSVP Summary Risk Ratings***

	Score		
Rating	**0**	**1**	**2**
Case Prioritization	46	41	13
Serious Harm	79	18	3
Immediate Action	84	9	7

Note. $N = 100$. 0 = *Low/routine/no*; 1 = *Moderate/elevated/possible*; 2 = *High/urgent/yes*.

Table 13 summarizes the Pearson correlations (r) between the RSVP Presence and Relevance total and domain scores. The correlation between Presence and Relevance total scores was very high, $r = .95$, $p < .001$; the same was true of the correlations between corresponding Presence and Relevance domain scores, $.86 \leq r \leq .95$, all $p < .001$ ($Mdn = .92$). Most of the correlations among non-corresponding domain scores, both within and between the Presence and Relevance ratings, were moderate to large in magnitude ($r \approx .30$ to $r \approx .60$), and statistically significant ($p < .001$). The exception was Sexual violence history domain, especially for Presence ratings, which

tended to have smaller correlations with scores on other domains—in some cases, the correlations were small and non-significant.

Table 13 Correlation (r) Between RSVP Presence and Relevance Ratings, Domains and Total

	Presence						Relevance					
	SEX (A)	**PSY (B)**	**MEN (C)**	**SOC (D)**	**MAN (E)**	**TOT (F)**	**SEX (G)**	**PSY (H)**	**MEN (I)**	**SOC (J)**	**MAN (K)**	**TOT (L)**
B	.25*											
C	.37***	.43***										
D	.14	.50***	.58***									
E	.17	.42***	.44***	.51***								
F	.61***	.70***	.80***	.74***	.68***							
G	.94***	.28**	.41***	.19	.27**	64***						
H	.28**	.87***	.45***	.55***	.47***	.71***	.33***					
I	.34***	.44***	.92***	.56***	.44***	.77***	.39***	.45***				
J	.21*	.50***	.54***	.86***	.47***	.71***	.28**	.58***	.55***			
K	.16	.42***	.43***	.52***	.95***	.67***	.27**	.49***	.46***	.53***		
L	.59***	.66***	.75***	.69***	.68***	.95***	.67***	.75***	.77***	.77***	.72***	

Note. N = 100. SEX = History of sexual violence, PSY = Psychological adjustment, MEN = Mental disorder, SOC = Social adjustment, MAN = Manageability, TOT = Total scores.

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

Table 14 presents the correlations between the RSVP Presence and Relevance total scores and the Summary Risk ratings. The Presence and Relevance total scores had correlations with the Summary Risk ratings that were moderate to large in magnitude and statistically significant, $.40 \leq r \leq .63$, all $p < .001$. The correlations between the Summary Risk ratings were smaller in magnitude albeit still statistically significant, $.20 \leq r \leq .63$, all $p < .05$. The smaller correlations for the Summary Risk Ratings was no doubt due in part to the fact they are trichotomous in nature.

Table 14 ***Correlation (r) Between RSVP Presence, Relevance, and Summary Risk Ratings***

	Presence Total	Relevance Total	Case Prioritization	Serious Harm
Relevance Total	.95***			
Case Prioritization	.63***	.63***		
Serious Harm	.43***	.43***	.32**	
Immediate Action	.40***	.41***	.63***	.20*

Note. N = 100.
*p < .05, **p < .01., *** p < .001

Interrater Reliability

Overall, the interrater reliability for the RSVP was good to excellent. The results for the Presence and Relevance ratings of individual risk factors are presented in Table 15. For Presence ratings, ICC_1 ranged from .58 to. 94 (M = .77, *Mdn* = .78), and ICC_2 ranged from .67 to .96 (M = .85, *Mdn* = .86). For the Relevance risk factors, ICC_1 ranged from .48 to .92 (M = .74, *Mdn* = .74), and ICC_2 ranged from .65 to .96 (M = .85, *Mdn* = .85).

The results for Presence and Relevance total and domain scores are presented in Table 16. Interrater reliability was excellent for both Presence total scores (ICC_1 = .93, ICC_2 = .96) and Relevance total scores (ICC_1 = .90, ICC_2 = .95). Interrater was slightly lower, but still good to excellent, for Presence domain scores (ICC_1 =.75 to .92, ICC_2 = .86 to .96) and Relevance domain scores ($ICC1_1$ = .75 to .89, ICC_2 = .86 to .94).

Table 15 Interrater Reliability of RSVP Presence and Relevance Ratings, Individual Risk Factors

	Presence		Relevance	
RSVP Risk Factors	**ICC_1**	**ICC_2**	**ICC_1**	**ICC_2**
Chronicity of sexual violence	.91	.96	.86	.92
Diversity of sexual violence	.68	.81	.69	.81
Escalation of sexual violence	.62	.77	.67	.80
Physical coercion	.89	.94	.86	.93
Psychological coercion	.83	.91	.85	.92
Extreme minimization	.70	.82	.70	.83
Attitudes	.68	.81	.71	.83
Problems with self-awareness	.61	.76	.48	.65
Problems with stress or coping	.59	.74	.58	.74
Problems from child abuse	.87	.93	.78	.87
Sexual deviance	.81	.90	.83	.91
Psychopathic personality disorder	.76	.86	.76	.86
Major mental illness	.84	.92	.83	.91
Problems with substance use	.94	.97	.85	.92
Violent or suicidal ideation	.77	.87	.64	.78
Intimate relationships	.72	.84	.69	.82
Non-intimate relationships	.58	.73	.67	.80
Problems with employment	.78	.88	.83	.90
Nonsexual criminality	.78	.88	.64	.78
Problems with planning	.69	.82	.63	.78
Problems with treatment	.87	.93	.77	.87
Problems with supervision	.91	.96	.92	.96

Note. $N = 100$.

Table 16 ***Interrater Reliability of RSVP Presence and Relevance Ratings, Domain and Total Scores***

	Presence		Relevance	
Scores	**ICC_1**	**ICC_2**	**ICC_1**	**ICC_2**
History of sexual violence	.90	.95	.89	.94
Psychological adjustment	.75	.86	.75	.86
Social adjustment	.82	.90	.82	.90
Manageability	.92	.96	.87	.93
Mental disorder	.87	.93	.80	.89
Total	.93	.96	.90	.95

Note. $N = 100$.

Finally, the interrater reliability of the RSVP Summary Risk ratings was also excellent: for Case Prioritization, $ICC_1 = .74$ and $ICC_2 = .85$; for Serious Harm, $ICC_1 = .85$ and $ICC_2 = .92$; and for Immediate Action, $ICC_1 = .80$ and $ICC_2 = .92$. These findings are presented in Tables 17, 18, and 19. The results are quite impressive, given the response formats for the ratings are trichotomous in nature and few participants received high scores on any of the them.

Table 17 ***Interrater Reliability of RSVP Case Prioritization Ratings***

	Rater 2			
Rater 1	**0**	**1**	**2**	**Total**
0	42	10	0	52
1	7	28	4	39
2	0	2	7	9
Total	49	40	11	100

Note. N = 100. 0 = *Low/routine*; 1 = *Moderate/elevated*; 2 = *High/urgent.* ICC_1= .74, ICC_2 = .85.

Table 18 ***Interrater Reliability of RSVP Serious Harm Ratings***

	Rater 2			
Rater 1	**0**	**1**	**2**	**Total**
0	75	2	0	77
1	5	14	0	19
2	0	1	3	4
Total	80	17	3	100

Note. N = 100. 0 = *Low*; 1 = *Moderate*; 2 = *High.* ICC_1= .85, ICC_2 = .92.

Table 19 ***Interrater Reliability of RSVP Immediate Action Ratings***

	Rater 2			
Rater 1	**0**	**1**	**2**	**Total**
0	82	1	1	84
1	2	6	4	12
2	0	1	3	4
Total	84	8	8	100

Note. N = 100. 0 = *No*; 1 = *Possible*; 2 = *Yes.* ICC_1= .80, ICC_2 = .89.

Concurrent Validity: SVR-20

Table 20 presents the correlations of the RSVP Presence and Relevance total and domain scores and Summary Risk ratings with the SVR-20 ratings. The correlations between corresponding RSVP and SVR-20 domain and total scores were all very large in magnitude and statistically significant, $.71 \leq r \leq .93$, all $p < .001$. (Recall that the RSVP Mental disorder domain corresponds to the SVR-20 Psychological adjustment domain; the RSVP Psychological adjustment domain has no parallel in the SVR-20.) Looking at the Summary Risk ratings, the correlations between Case Prioritization ratings and the SVR-20 total and domain scores were moderate to large in magnitude and statistically significant, $.40 \leq r \leq .66$, all $p < .001$. The correlations between Serious Harm and Immediate Action ratings and SVR-20 total and domain scores were small to moderate in magnitude, but still statistically significant, $.22 \leq r \leq .66$, all $p < .05$.

In terms of individual items, given the significant overlap between the RSVP and SVR-20, concurrent validity between the individual items was not evaluated.

Table 20 Correlation (r) Between RSVP Presence, Relevance, and Summary Risk Ratings and SVR-20 Ratings

	SVR-20 Ratings				
RSVP Ratings	**Total**	**SEX**	**PSY**	**SOC**	**FUT**
Presence					
SEX	.51***	.80***	.35***	.12	.16
PSY	.67***	.46***	.57***	.51***	.44***
MEN	.81***	.42***	.92***	.58***	.34**
SOC	.74***	.24*	.61***	.84***	.45***
MAN	.71***	.38***	.41***	.70***	.77***
Total	.97***	.69***	.81***	.74***	.58***
Relevance					
SEX	.57***	.81***	.39***	.19	.23
PSY	.70***	.54***	.52***	.52***	.49***
MEN	.77***	.44***	.85***	.52***	.37***
SOC	.70***	.36***	.52***	.71***	.49***
MAN	.70***	.38***	.39***	.65***	.82***
Total	.93***	.73***	.73***	.67***	.63***
Summary Risk					
Case Prioritization	.66***	.56***	.52***	.40***	.46***
Serious Harm	.46***	.35***	.40***	.34***	.24*
Immediate Action	.42***	.43***	.32**	.22*	.25*

Note. *N* = 100. SEX = History of sexual violence/offences, PSY = Psychological adjustment, MEN = Mental disorder, SOC = Social adjustment, MAN = Manageability, FUT = Future Plans.

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

Concurrent Validity: Static-99R, Static-2002R, and SORAG

The correlations of RSVP Presence and Relevance total and domain scores and Summary Risk ratings with total scores and risk categories of the actuarial risk assessment instruments are presented in Table 21. Focusing on the RSVP Presence and Relevance ratings, correlations between Presence total scores and actuarial risk assessment instrument total scores and risk categories were moderate to large in magnitude and statistically significant, $.38 \leq r \leq .68$, all $p < .001$. The same was true for the correlations between RSVP Relevance total scores and actuarial risk assessment instrument total scores and risk categories, $.40 \leq r \leq .66$, all $p < .001$. Of the Presence and Relevance domains, the largest correlations were between Manageability scores and actuarial risk assessment instrument total scores and risk categories, and the smallest (some of which were not statistically significant) were between Sexual violence history scores and actuarial risk assessment instrument total scores and risk categories.

Focusing next on RSVP Summary Risk ratings, correlations between Case Prioritization ratings and actuarial risk assessment instrument total scores and risk categories were large in magnitude and statistically significant, $.63 \leq r \leq .67$, all $p < .001$. Correlations between Serious Harm and Immediate Action ratings and actuarial risk assessment instrument total scores and risk categories were variable but generally smaller in magnitude, $.14 \leq r \leq .54$, and some were not statistically significant.

For the sake of completeness, the concurrent validity of the RSVP individual risk factors with respect to Static-99R, Static-2002R, and SORAG total scores and risk categories is presented in Appendix G, Tables G7 and G8.

Table 21 ***Correlation (r) Between RSVP Presence, Relevance, and Summary Risk Ratings and Static-99R, Static-2002R, and SORAG Total and Risk Categories***

	Static99R		Static2002R		SORAG	
RSVP	**Total**	**Risk Categories**	**Total**	**Risk Categories**	**Total**	**Bins**
Presence						
SEX	.06	.09	.09	.10	.15	.16
PSY	.26**	.24*	.27**	.21*	.49***	.48***
MEN	.38***	.36***	.41***	.40***	.53***	.55***
SOC	.30**	.20*	.28**	.19	.62***	.60***
MAN	.52***	.50***	.59***	.57***	.70***	.67***
Total	.41***	.38***	.45***	.40***	.68***	.67***
Relevance						
SEX	.17	.19	.21*	.21*	.25*	.26**
PSY	.27**	.27**	.25*	.21*	.51***	.52***
MEN	.39***	.37***	.40***	.39***	.51***	.52***
SOC	.27**	.18	.26*	.15	.55***	.53***
MAN	.51***	.50***	.56***	.53***	.68***	.65***
Total	.43***	.40***	.45***	.40***	.66***	.65***
Summary Risk						
Case Prioritization	.63***	.66***	.67***	.66***	.65***	.66***
Serious Harm	.24*	.27**	.14	.27**	.44***	.43***
Immediate Action	.50***	.54***	.52***	.52***	.44***	.47***

Note. *N* = 100. SEX = History of sexual violence/offences, PSY = Psychological adjustment, MEN = Mental disorder, SOC = Social adjustment, MAN = Manageability.

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

Predictive Validity: Risk Categories

For descriptive purposes, I calculated the recidivism rate for participants in each of the risk categories for the various risk measures.

Table 22 presents the proportion of recidivists for RSVP Case Prioritization. The proportion of recidivists increased steadily with Case Prioritization ratings. Using the *Low* group as the reference, the increased risk (odds) for recidivism—as indexed using the Odds Ration, or OR—was 2.20 times greater in the *Moderate* group and 7.60 times higher in the *High* group. The overall difference among the proportion of recidivists in the three groups was significant, χ^2 (2, N = 100) = 9.74, p = .008.

Table 22 ***Proportion of Recidivists in RSVP Case Prioritization Categories***

	Total number	Proportion recidivists	95% CI	OR	χ^2 (1)	*p*
Low	46	.17	[.09, 31]	1.00		
Moderate	41	.32	[.19, .48]	2.20	2.40	.122
High	13	.62	[.33, .84]	7.60	9.83	.002

Note. Overall χ^2 (2, N = 100) = 9.74, p = .008.

Table 23 presents the proportion of recidivists for Static-99R risk categories. The proportion of recidivists increased steadily across the four categories. Using the *Low* group as the reference, the increased was 2.25 times greater in the *Low-moderate* group, 2.47 times greater in the *Moderate-high* group, and 9.00 times higher in the *High* group. The overall difference among the proportion of recidivists in the four categories was significant, χ^2 (3, N = 100) = 9.89, p = .020.

Table 23 Proportion of Recidivists in Static-99R Risk Categories

	Total number	Proportion recidivists	95% CI	OR	χ^2 (1)	*p*
1	28	.14	[.05, .33]	1.00		
2	33	.27	[.15, .45]	2.25	1.50	.220
3	24	.29	[.14, .51]	2.47	1.68	.195
4	15	.60	[.34, .82]	9.0	9.45	.002

Note. Overall χ^2 (3, N = 100) = 9.89, p = .020.

Table 24 presents the proportion of recidivists for Static-2002R risk categories. The proportion of recidivists increased steadily across the five categories. Using the *Low* group as the reference, the increase was 1.83 times greater in the *Low-moderate* group, 3.30 times greater in the *Moderate* group, 7.33 times higher in the *Moderate-high* group, and 14.67 times higher in the *High* group. The overall difference among the proportion of recidivists in the five categories was significant, χ^2 (4, N = 100) = 13.19, p = .010.

Table 24 Proportion of Recidivists in Static-2002R Risk Categories

	Total number	Proportion recidivists	95% CI	OR	χ^2 (1)	*p*
1	25	.12	[.04, .32]	1.00		
2	25	.20	[.08, .41]	1.83	.58	.445
3	29	.31	[.17, .50]	3.30	2.76	.097
4	12	.50	[.23, .77]	7.33	6.19	.013
5	9	.67	[.31, .90]	14.67	9.86	.002

Note. Overall χ^2 (4, N = 100) = 13.19, p = .010.

Finally, Table 25 presents the proportion of recidivists for SORAG bins (risk categories). The proportion of recidivists tended to increase across the 9 bins, but in a very inconsistent manner. Using Bin 1 as a reference group, the risk actually decreased in Bins 2, 3, and 4 (OR = 0.25, 0.79, and 0.87, respectively); increased in Bin 5 (OR = 5.30); decreased again in Bins 6 and 7 (both OR = 4.33); increased in Bin 8 (OR = 6.50); and then decreased in Bin 9 (OR = 1.44). Notwithstanding this inconsistency, the overall

difference among the proportion of recidivists in the 9 bins was significant, χ^2 (8, N = 100) = 18.99, p = .015.

Table 25 Proportion of Recidivists in SORAG Risk Categories

	Total number	Proportion recidivists	95%CI	OR	χ^2 (1)	*p*
1	16	.19	[.06, .46]	1.00		
2	18	.06	[.01, .32]	.25	1.38	.240
3	13	.15	[.04, .47]	.79	.05	.815
4	12	.17	[.04, .50]	.87	.02	.889
5	20	.55	[.33, .75]	5.30	4.78	.029
6	4	.50	[.09, .91]	4.33	1.58	.208
7	8	.50	[.18, .82]	4.33	2.42	.120
8	5	.60	[.17, .92]	6.50	3.03	.082
9	4	.25	[.02, .82]	1.44	.07	.785

Note. Overall χ^2 (8, N = 100) = 18.99, p = .015.

Overall, the ability of the RSVP to distinguish between recidivists and non-recidivists using a risk category (i.e., Case Prioritization ratings) was more or less equivalent to that of the other risk measures.

Predictive Validity: AUCs

Next, I conducted ROC analyses, as AUC is a commonly used method of indexing predictive validity. AUC values for RSVP Presence and Relevance ratings for the individual risk factors are presented in Table 26. Looking first at Presence ratings, 2 out of the 22 items had significant predictive validity: Problems with supervision and Psychopathic personality disorder, AUC = .71 and .63, respectively, both $p \leq .05$. Looking next at Relevance ratings, 5 of the 22 items had significant predictive validity ($p \leq .05$): Problems with supervision, AUC = .71; Problems with treatment, AUC = .65; Diversity of sexual violence, AUC = .64; Sexual deviance, AUC = .63; and Psychopathic personality disorder, AUC = .63.

For the sake of completeness, the AUC values for the individual risk factors or items of the SVR-20, Static-99R, Static-2002R, and SORAG are presented in Appendix G, Tables G9 to G12. For the SVR-20, 2 out of 20 items significantly predicted sexual recidivism: Past supervision failure, AUC = .68, $p < .01$; and Psychopathy, AUC = .63, $p < .05$. For the Static-99R, the only item to significantly predict sexual recidivism was Prior sex offences, AUC = .63, $p < .05$. For the Static-2002R, the only item to significantly predict sexual recidivism was Persistence of sexual offending, AUC = .63, $p < .05$. For the SORAG, two of 13 items significantly predicted sexual recidivism: Failure on prior conditional release, AUC = .70, $p < .01$; and Elementary school maladjustment, AUC = .64, $p < .05$.

Table 26 ***AUC for RSVP Presence and Relevance Ratings, Individual Risk Factors***

	Presence			Relevance		
Risk Factor	**AUC**	**95% CI**	***p***	**AUC**	**95% CI**	***p***
Chronicity of sexual violence	.51	[.38, .63]	.933	.55	[.42, .67]	.457
Diversity of sexual violence	.62	[.50, .75]	.054	**.64**	[.51, .76]	.030
Escalation of sexual violence	.51	[.38, .63]	.939	.53	[.41, .66]	.592
Physical coercion in sexual violence	.59	[.47, .72]	.149	.59	[.47, .72]	.149
Psychological coercion in sexual violence	.43	[.31, .56]	.305	.46	[.33, .58]	.494
Extreme minimization/denial of sexual violence	.47	[.34, .59]	.608	.49	[.36, .61]	.820
Attitudes that support sexual violence	.59	[.47, .71]	.166	.61	[.49, .72]	.101
Problems with self-awareness	.53	[.41, .66]	.600	.53	[.40, .65]	.682
Problems with stress or coping	.57	[.45, .69]	.286	.56	[.44, .68]	.346
Problems resulting from child abuse	.56	[.44, .68]	.368	.58	[.46, .70]	.230
Sexual deviance	.61	[.49, .74]	.075	**.63**	[.51, .75]	.047
Psychopathic personality disorder	**.63**	[.51, .76]	.040	**.63**	[.51, .76]	.040
Major mental illness	.57	[.44, .69]	.305	.53	[.41, .65]	.624
Problems with substance use	.51	[.39, .64]	.855	.49	[.37, .62]	.891
Violent or suicidal ideation	.59	[.47, .72]	.153	.57	[.44, .69	.309
Problems with intimate relationships	.54	[.42, .66]	.514	.52	[.40, .65]	.738
Problems with non-intimate relationships	.55	[.43, .67]	.412	.57	[.45, .69]	.288
Problems with employment	.55	[.42, .68]	.434	.53	[.41, .65]	.668
Nonsexual criminality	.60	[.47, .72]	.138	.54	[.41, .66]	.569
Problems with planning	.61	[.50, .73]	.077	.62	[.51, .74]	.058
Problems with treatment	.60	[.48, .72]	.118	**.65**	[.53, .77]	.019
Problems with supervision	**.71**	[.60, .82]	.001	**.71**	[.59, .82]	.001

Note. $N = 100$. AUCs in boldface type were statistically significant, $p < .05$.

The AUC values for RSVP Presence and Relevance domain scores are presented in Table 27. AUC was moderate to large in magnitude and statistically significant for the Manageability domain, AUC = .71 for Presence and AUC = .73 for Relevance, both $p < .05$. AUC was small to moderate in magnitude and statistically significant for the Mental disorder domain, AUC = .66 for Presence and AUC = .65 for Relevance, both $p < .05$.

Table 27 AUC for RSVP Presence and Relevance Ratings, Domains

	Presence			Relevance		
Ratings	**AUC**	**95%CI**	***p***	**AUC**	**95%CI**	***p***
History of sexual violence	.56	[.43, .69]	.352	.60	[.47, .72]	.139
Psychological adjustment	.60	[.47, .72]	.130	.61	[.49, .73]	.098
Mental disorder	**.66**	[.54, .78]	.012	**.65**	[.53, .77]	.021
Social adjustment	.58	[.47, .70]	.198	.55	[.43, .66]	.471
Manageability	**.71**	[.59, .82]	.001	**.73**	[.61, .84]	< .001

Note. $N = 100$. AUCs in boldface type were statistically significant, $p < .05$.

The AUC values for the total scores and risk categories for all the risk measures are presented in Table 28. They were very similar: all were moderate to large in magnitude and statistically significant, $.67 \leq AUC < .71$, all $p \leq .01$. Overall, the predictive validity of judgements made using the RSVP appeared equivalent to that of the other risk measures. This was confirmed by χ^2 tests of the homogeneity of the AUCs for total scores and for risk categories, neither of which was significant (see Table 29).

Table 28 AUC for All Risk Measures, Total and Risk Categories

	AUC	95%CI	*p*
RSVP			
Total	**.68**	[.57, .80]	.004
Case Prioritization	**.67**	[.56, .78]	.010
SVR-20			
Total	**.68**	[.56, .80]	.006
Static-99R			
Total	**.68**	[.56, .80]	.015
Risk Categories	**.67**	[.56, .78]	.010
Static-2002R			
Total	**.71**	[.59, .82]	.002
Risk Categories	**.71**	[.60, .82]	.005
SORAG			
Total	**.71**	[.59, .82]	.001
Bins	**.71**	[.60, .82]	.001

Note. *N* = 100. AUCs in boldface type were statistically significant, $p < .05$.

Table 29 Comparison of AUCs for Risk Measures, Total and Risk Categories

Comparison	χ^2	*df*	*p*
AUCs for Total Scores	2.00	4	.736
AUCs for Risk Categories	2.56	4	.464

Note. *N* = 100.

Predictive Validity: Survival Analyses

First, I conducted survival analysis using the Kaplan-Meier method to examine the overall survivor function, which is illustrated in Figure 1. Recall that there were 100 participants, of whom 29 were recidivists (failures). The overall time at risk for the sample was 257,182 days. The *Mdn* time at risk per participant was 3,275 days (M = 2,572 days, range = 33 to 3592 days). Q1 (25%ile) for survival time was 1,750 days.

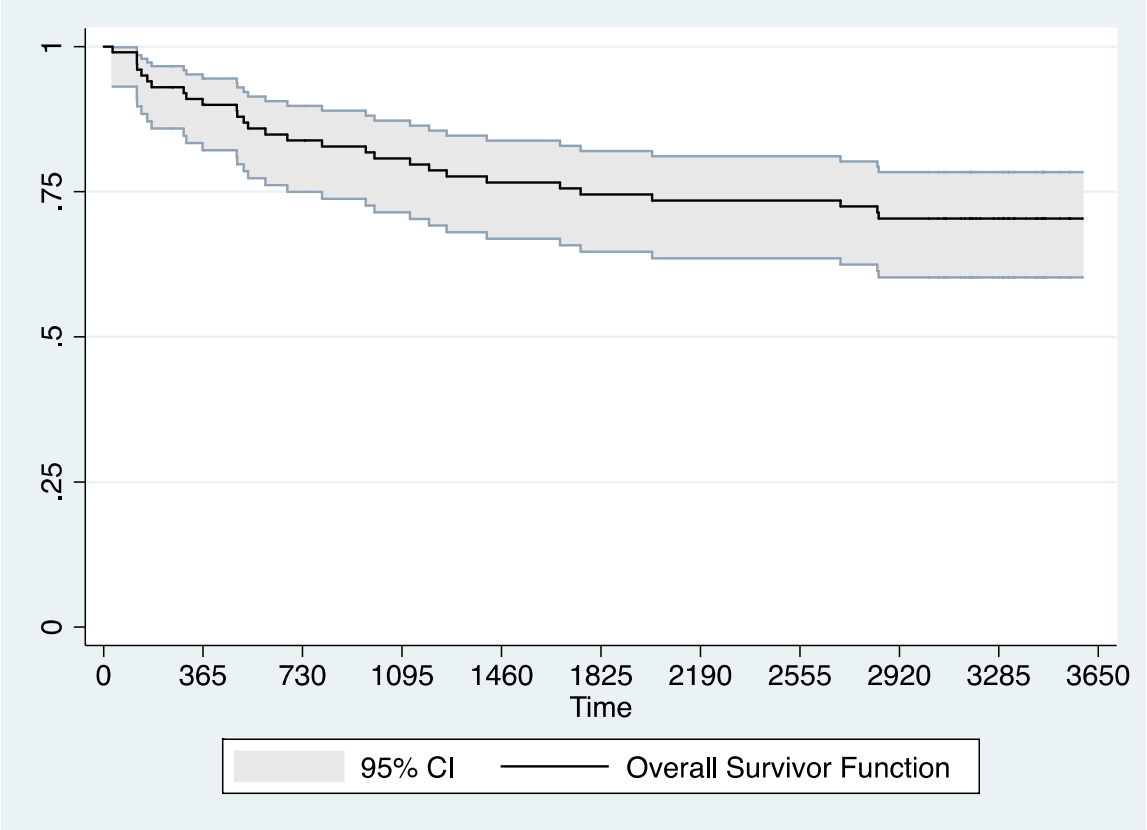

Figure 1 Kaplan-Meier Survival Analysis: Overall Survivor Function

I then conducted Kaplan-Meier survival analysis to evaluate the predictive validity of RSVP Case Prioritization ratings. The survival curves are presented in Figure 2. As the figure reveals, the likelihood and speed of failure was associated with the Case Prioritization ratings. The overall log-rank test indicated that the survivor functions were not equivalent, χ^2 (2, *N* = 100) = 12.28, *p* = .002. The log-rank test of trend of the survivor function across ordered groups was also significant, χ^2 (1, *N* = 100) = 10.83, *p* = .001.

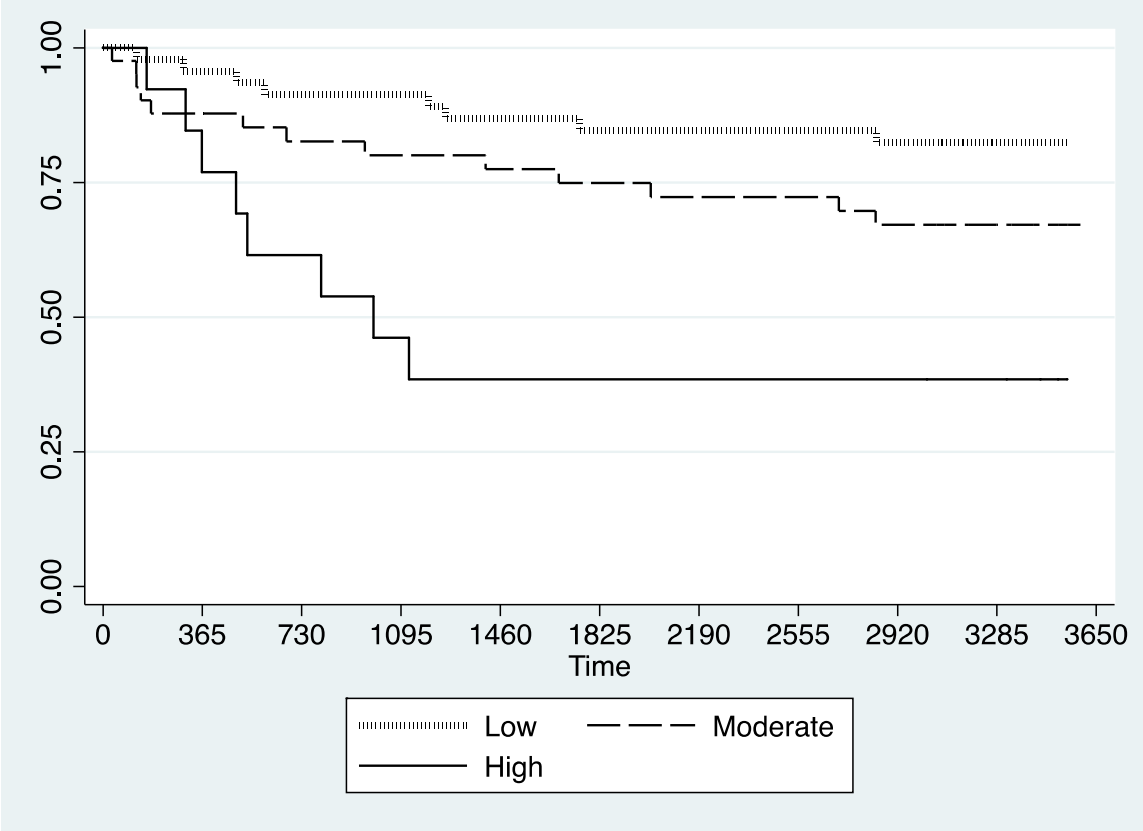

Figure 2 Kaplan-Meier Survival Functions, RSVP Case Prioritization Ratings

Next, I evaluated the predictive validity of the risk measures using Cox regression survival analysis. This method of survival analysis permits modelling of the predictive validity of multiple independent variables simultaneously, both in terms of overall model fit via a χ^2 test and the predictive power of independent variables via examination of Hazard Ratios (HRs). Note that Cox regression survival analysis assumes the HRs for groups are proportional and constant over time; this is sometimes referred to as the proportional hazards assumption. This assumption is tested in Stata by examining the linear regression of the scaled Schoenfeld residuals on time and determining, for individual independent variables and globally, whether the slope of the Schoenfeld residuals deviates significantly from zero. I tested the proportional hazards assumption for all of the analyses discussed below, and in each case the test was non-significant, $p > .05$. Note also that Cox regression survival analysis may also be adversely impacted by collinearity. There are actually several forms of collinearity that may cause problems, but most of them result in major problems when attempting to fit a model, such as “failure to converge” errors, extremely large HRs, or HRs with missing

SE estimates. No such errors were observed in the analyses discussed below. Another form of collinearity, collinearity among independent variables, may not result in such obvious problems but can greatly distort the interpretation of HRs. I directly evaluated collinearity among independent variables using the Variance Inflation Factor, or VIF. Typically, VIF values between 1 and 4 are interpreted as reflecting negligible or small collinearity; between 4 and 10, as reflecting substantial collinearity; and 10 or greater, as reflecting very large and potentially problematic collinearity (e.g., O'Brien, 2007).

I started by conducting a series of Cox regressions examining the predictive power of the RSVP Presence and Relevance ratings for the individual risk factors. These are presented in Table 30. As the purpose of the analysis was primarily descriptive and the risk factors were evaluated one at a time, the table presents only the HRs along with their 95%CI and *p* level (rather than this information plus information concerning overall model fit). Looking first at the Presence ratings, 20 out of 22 risk factors had HRs that were positive, and the *Mdn* HR was 1.38; 3 out of 22 HRs were positive and statistically significant ($p < .05$). Looking next at the Relevance ratings, 6 out of 22 items were statistically significant; Diversity of sexual violence, Problems with supervision (both $p < .01$), Sexual deviance, Psychopathic personality disorder, Problems with planning, and Problems with treatment (all $p < .05$ level). Again, 20 out of 22 risk factors had HRs that were positive, and the *Mdn* HR was 1.38; but this time 6 out of 22 HRs were positive and statistically significant ($p < .05$). Across both analyses, the risk factors that were most strongly associated with the risk for recidivism were Diversity of sexual violence, Psychopathic personality disorder, and Problems with supervision.

Table 30 ***Summary of Cox Regression Analyses of RSVP Presence Ratings for Individual Risk Factors***

	Presence			Relevance		
Risk Factor	**HR**	**95%CI**	***p***	**HR**	**95%CI**	***p***
Chronicity of sexual violence	1.05	[.69, 1.61]	.814	1.21	[.79, 1.86]	.391
Diversity of sexual violence	**1.77**	[1.11, 2.82]	.016	**1.78**	[1.18, 2.69]	.006
Escalation of sexual violence	1.04	[.67, 1.60]	.875	1.12	[.72, 1.74]	.604
Physical coercion in sexual violence	1.32	[.89, 1.95]	.170	1.31	[.87, 1.95]	.197
Psychological coercion in sexual violence	.81	[.53, 1.22]	.310	.48	[.56, 1.31]	.477
Extreme minimization/denial of sexual violence	.95	[.56, 1.59]	.833	.99	[.61, 1.62]	.979
Attitudes that support sexual violence	1.47	[.84, 2.56]	.177	1.58	[.93, 2.69]	.093
Problems with self-awareness	1.34	[.42, 4.25]	.622	1.17	[.46, 2.98]	.745
Problems with stress or coping	1.66	[.75, 3.69]	.210	1.56	[.75, 3.26]	.234
Problems resulting from child abuse	1.26	[.84, 1.89]	.260	1.34	[.88, 2.04]	.174
Sexual deviance	1.54	[.99, 2.40]	.055	**1.62**	[1.04, 2.52]	.032
Psychopathic personality disorder	**1.67**	[1.07, 2.61]	.025	**1.67**	[1.07, 2.61]	.025
Major mental illness	1.28	[.83, 1.98]	.263	1.14	[.72, 1.82]	.574
Problems with substance use	1.07	[.70, 1.62]	.770	1.00	[.67, 1.49]	.993
Violent or suicidal ideation	1.41	[.93, 2.15]	.108	1.28	[.82, 1.99]	.271
Problems with intimate relationships	1.32	[.55, 3.16]	.536	1.10	[.47, 2.57]	.824
Problems with non-intimate relationships	1.69	[.66, 4.33]	.278	1.59	[.76, 3.34]	.220
Problems with employment	1.28	[.76, 2.13]	.351	1.16	[.75, 1.81]	.501
Nonsexual criminality	1.42	[.90, 2.24]	.130	1.13	[.71, 1.78]	.614
Problems with planning	2.71	[.98, 7.49]	.054	**2.81**	[1.05, 7.56]	.040
Problems with treatment	1.72	[.94, 3.14]	.080	**2.24**	[1.15, 4.38]	.018
Problems with supervision	**2.35**	[1.43, 3.88]	.001	**2.22**	[1.43, 3.44]	<.001

Note. HR = Hazard Ratio. HRs in boldface type were statistically significant, $p < .05$.

Next, I evaluated the predictive validity of RSVP and compared it to that of the other risk measures. As the magnitude of a HR reflects the increase in relative risk associated with a 1-unit increase in the value of the predictor variable, the magnitude of HRs is directly comparable only when the predictor variables have the same metric. For this reasons, I recoded total scores on each risk measure into five (approximately) equal-sized groups. Table 31 presents the number of participants in each of the five groups for the various risk assessment measures. In the analyses that follow, I focused solely on RSVP Presence scores, as the analyses of Relevance scores yielded virtually identical results.

Table 31 Distribution of Participants into Five Groups for RSVP Presence, SVR-20, Static-99R, Static-2002R and SORAG Total Scores

	Frequency/Percentage				
Group	**RSVP Presence**	**SVR-20**	**Static-99R**	**Static-2002R**	**SORAG**
1	16	17	14	16	19
2	20	18	14	17	20
3	20	24	15	17	20
4	21	14	33	29	21
5	23	27	24	21	20

Note. $N = 100$.

I started by conducting descriptive analyses to examine the predictive accuracy of the various risk measures, both total scores (recoded into five equal-sized groups) and risk categories. As evident in Table 32, all of the predictors performed similarly and all were significantly related to sexual recidivism ($p < .01$).

Table 32 Summary of Cox Regression Analyses of Risk Measures

Risk Measure	HR	95%CI	*p*
RSVP			
Five Groups	**1.61**	[1.19, 2.16]	.002
Case Prioritization	**2.27**	[1.37, 3.75]	.001
SVR-20			
Five Groups	**1.50**	[1.14, 1.99]	.004
Static-99R			
Five Groups	**1.56**	[1.12, 2.17]	.008
Risk Categories	**1.20**	[1.06, 1.37]	.005
Static-2002R			
Five Groups	**1.67**	[1.21, 2.30]	.002
Risk Categories	**1.25**	[1.10, 1.43]	.001
SORAG			
Five Groups	**1.66**	[1.24, 2.21]	.001
Bins	**1.04**	[1.02, 1.07]	.002

Note. HR = Hazard Ratio. HRs in boldface type were statistically significant, $p < .01$.

I then examined the potential confounding influence of age at the end of treatment. To do this, I recoded age into 5 equal sized groups. The number of people in each group was as follows: 1, $n = 20$; 2, $n = 20$; 3, $n = 18$; 4, $n = 21$; and 5, $n = 21$. Age did not have significant predictive validity on its own, HR = .83, 95%CI [.64, 1.08], $p = .164$, but when entered into predictive models for the risk measures it tended to significantly improve fit. I therefore decided to control for age (five groups) in subsequent analyses.

Next, I evaluated the predictive accuracy of judgements made using the RSVP vis-à-vis those made using other risk measures, controlling for age. I did this using hierarchical Cox regression analyses. I started by entering age and the RSVP ratings on Block One, and then entered one of the other risk measures on Block Two. The key test was whether addition of the other risk measures significantly increased model fit (i.e., predictive validity). The findings for analyses of the Five Group total scores are

presented in Table 33. As the table indicates, only the incremental validity for Static-2002R total scores was significant, $p < .05$. The VIF values were large (> 7) for the analysis involving RSVP and SVR-20 total scores, however, suggesting its results should be interpreted with caution due to substantial collinearity.

The findings for analyses of the RSVP Case Prioritization ratings vis-à-vis the other risk categories are presented in Table 34. None of the incremental validities were statistically significant.

Overall, these results suggest that the predictive validity of the RSVP was equivalent to that of the other risk measures.

Table 33 ***Cox Regression Analyses Evaluating Incremental Validity of Other Risk Measures (Block Two) vis-à-vis the RSVP and Age (Block One), Five Groups***

Variable	HR	95%CI	*p*	VIF
Block One: χ^2 (2, *N* = 100) = 13.02, *p* = .002				
Age Five Groups	.82	[.63, 1.58]	.154	1.00
RSVP Five Groups	**1.61**	[1.19, 2.16]	.002	1.00
Block Two: $\Delta\chi^2$ (1, *N* = 100) = 0.28, *p* = .597				
Age Five Groups	.81	[.61, 1.07]	.133	1.06
RSVP Five Groups	1.95	[.88, 4.31]	.100	7.63
SVR-20 Five Groups	.82	[.38, 1.75]	.600	7.65
Block Two: $\Delta\chi^2$ (1, *N* = 100) = 1.68, *p* = .194				
Age Five Groups	.92	[.63, 1.58]	.579	1.42
RSVP Five Groups	**1.48**	[1.06, 1.98]	.017	1.18
Static-99R Five Groups	1.27	[.87, 1.85]	.208	1.58
Block Two: $\Delta\chi^2$ (1, *N* = 100) = 3.96, *p* = .047				
Age Five Groups	.91	[.68, 1.20]	.494	1.23
RSVP Five Groups	**1.42**	[1.04, 1.94]	.027	1.19
Static-2002R Five Groups	1.41	[.99, 1.99]	.054	1.41
Block Two: $\Delta\chi^2$ (1, *N* = 100) = 2.10, *p* = .147				
Age Five Groups	.82	[.67, 1.22]	.512	1.33
RSVP Five Groups	1.34	[.90, 1.99]	.150	2.05
SORAG Five Groups	1.33	[.89, 1.99]	.161	2.34

Note. HR = Hazard Ratio. HRs in boldface type were statistically significant, $p < .05$. VIF = Variance Inflation Factor.

Table 34 ***Cox Regression Analyses Evaluating Incremental Validity of Other Risk Measures (Block Two) vis-à-vis the RSVP and Age (Block One), Risk Categories***

Variable	HR	95%CI	*p*	VIF
Block One: χ^2 (2, *N* = 100) = 11.48, *p* = .003				
Age Five Groups	.84	[.65, 1.10]	.208	1.00
RSVP Case Prioritization	**2.24**	[1.35, 1.72]	.002	1.00
Block Two: $\Delta\chi^2$ (1, *N* = 100) = 0.44, *p* = .507				
Age Five Groups	.87	[.66, 1.16]	.358	1.29
RSVP Case Prioritization	1.87	[.90, 3.89]	.094	1.96
Static-99R Risk Categories	1.20	[.70, 2.06]	.505	2.26
Block Two: $\Delta\chi^2$ (1, *N* = 100) = 3.14, *p* = .076				
Age Five Groups	.92	[.69, 1.24]	.585	1.26
RSVP Case Prioritization	1.36	[.65, 2.88]	.415	2.07
Static-2002R Risk Categories	1.49	[.95, 2.35]	.085	2.34
Block Two: $\Delta\chi^2$ (1, *N* = 100) = 1.16, *p* = .281				
Age Five Groups	.89	[.67, 1.18]	.406	1.23
RSVP Case Prioritization	1.73	[.86, 3.48]	.125	1.95
SORAG Bins	1.12	[.91, 1.38]	.282	2.18

Note. HR = Hazard Ratio. HRs in boldface type were statistically significant, $p < .05$. VIF = Variance Inflation Factor.

DISCUSSION

Summary of Major Findings

Overall, the results indicated that judgements of risk made using the RSVP had interrater reliability, concurrent validity, and predictive validity that was at least moderate or good in absolute terms and, in relative terms, equivalent to that of other commonly-used risk measures.

Interrater reliability

The interrater reliability of the RSVP was excellent for individual risk factors, total scores and Summary Risk ratings. When interpreting the cross tabulations, the level of agreement was good to excellent for the Summary Risk ratings. The level of agreement for Serious Harm and Immediate Action was higher than for Case Prioritization. This finding may have been related to the nature of the sample in terms of risk for violence. In general, with a relatively lower risk sample, the majority will not be at high risk for serious harm or in need of urgent intervention. It is also likely to be more clearly apparent when an offender is at increased risk for seriously hurting someone or in need of urgent action so as to prevent reoffending. Further, the guidelines for coding these two Summary Risk ratings are quite specific whereas the determination of Case Prioritization is likely more ambiguous for assessors.

When comparing the interrater reliability of the RSVP to the other risk assessment instruments, overall, the RSVP was relatively equivalent, although there were minor differences. The SVR-20 had slightly higher interrater reliability for individual risk factors than the RSVP. This may have been related to the coding guidelines for the RSVP being somewhat broader which lends itself to more subjectivity. Regardless, interrater reliability for both the RSVP and SVR-20 was excellent on an individual risk level. For both the Static-99R and Static-2002R, the average interrater reliability for

individual risk factors was slightly higher than the RSVP's average interrater reliability for individual items. The SORAG's average interrater reliability for individual items was slightly lower than the RSVP's average interrater reliability for individual items. But overall, the RSVP performed as well as the other risk assessment instruments in terms of interrater reliability for individual items. The same was true for total scores; the interrater reliability was almost identical across instruments with all of the ICC's reflecting excellent interrater reliability. The interrater reliability for the risk categories was also excellent for the RSVP and the actuarial instruments, although it was higher for the latter. This difference is likely due to the fact that the risk categories for the actuarial instruments were based on computed total scores. In other words, a statistical program was used to add up the total scores and assign risk categories. In contrast, RSVP Case Prioritization ratings were based on the judgement of the raters, which has an element of subjectivity. Despite this element of subjectivity, the interrater reliability for RSVP Case Prioritization was excellent, which suggests that risk ratings based on structured professional judgement can be reliably rated between different assessors.

Concurrent validity

The RSVP showed good concurrent validity overall with the SVR-20, Static-99R, Static-2002R and the SORAG, particularly with regards to total scores and risk categories. In terms of individual items, given the significant overlap between the RSVP and SVR-20, concurrent validity between the individual items was not evaluated. Approximately half of the RSVP individual items were significantly correlated with the Static-99R, Static-2002R and SORAG total scores and risk categories, with the strongest concurrent validity appearing to be with the SORAG. These results may be related to the fact that similar to the RSVP, the SORAG includes risk factors that are psychosocial in nature, whereas the Static-99R and Static-2002R focus on static risk factors predominantly related to criminal history.

There was good concurrent validity between the RSVP and SVR-20 total and domain scores; each like score was significantly correlated between the RSVP and SVR-20. For example, RSVP and SVR-20 total scores were significantly correlated, RSVP and SVR-20 five group total scores were significantly correlated, RSVP and SVR-20 Sexual history domain scores were significantly correlated, and so on. There was

good concurrent validity between the RSVP total scores and the actuarial instruments' total scores and risk categories. Overall, there was good concurrent validity between the RSVP domain scores and the actuarial instruments' total score and risk categories with the exception of RSVP Sexual history domain. The correlations between the RSVP Presence Sexual history domain scores and the actuarial instruments' total scores and risk categories were all non-significant. However, the RSVP Relevance Sexual history domain scores were significantly correlated with the Static-2002R and SORAG total scores and risk categories but not the Static-99R total scores or risk categories. There was also some variability between the RSVP Social adjustment domain scores. All of the correlations between Presence Social adjustment domain scores and the actuarial instruments' total scores and risk categories were significant. Yet for the Relevance Social adjustment domain scores, the correlations with the Static-99R and Static-2002R risk categories were non-significant.

All of the RSVP Summary Risk ratings showed good concurrent validity with SVR-20 total and domain scores. Overall, the RSVP Summary Risk ratings also had good concurrent validity with the actuarial instruments' total scores and risk categories. The only exception was between RSVP Serious Harm and the Static-2002R total scores where the correlation was non-significant. The correlations between RSVP Case Prioritization and actuarial instruments' total scores and risk categories were the greatest, although did not differ in terms of significance (all $p < .001$) relative to the correlations between RSVP Immediate Action and actuarial instruments' total scores and risk categories.

Predictive validity

The results indicated that overall, the RSVP had good predictive validity for sexual recidivism in terms of Case Prioritization and total scores. The results were varied with respect to the RSVP individual risk factors, domain scores and other Summary Risk ratings. There were three RSVP Presence risk factors that were predictive of sexual recidivism across the different types of analyses (AUC and Cox regression). These risk factors were Diversity of sexual violence, Psychopathic personality disorder and Problems with supervision, with the latter appearing to be the strongest individual predictor for sexual recidivism. It is not surprising that these three risk factors were

predictive of sexual recidivism. If an offender has engaged in diverse sexual violence, then it suggests that there are multiple scenarios (i.e., types of victims and/or types of offending behaviour) where the offender may be vulnerable to sexually recidivate. Diversity for sexual violence is also captured under other sexual violence risk assessment instruments, such as the Static-99R and Static-2002R, where different types of victims and offending behaviour are coded for. Psychopathy has been well established in the literature as a risk factor for sexual recidivism (e.g., Hanson & Morton-Bourgon, 2005). By their nature, individuals who rate high on psychopathy generally possess personality traits, such as impulsivity and self-regulation problems, which place these individuals at higher risk for offending behaviour, including sexual violence. With regards to problems with supervision, this risk factor has also been found to significantly predict sexual recidivism in previous research (Dempster & Hart, 2002) and is included in both the Static-2002R and the SORAG.

The results for the SVR-20 were similar to the RSVP with two risk factors significantly predicting sexual recidivism, Psychopathy and Past supervision failure. For the Static-99R, the only item to significantly predict sexual recidivism was Prior sex offences. For the Static-2002R, the only item to significantly predict sexual recidivism was Persistence of sexual offending. For the SORAG, the two items that significantly predicted sexual recidivism were Elementary school maladjustment and Failure on prior conditional release.

For the RSVP domain scores (Presence and Relevance) Mental disorder and Manageability significantly predict sexual recidivism based on AUC's. The other RSVP domain scores were not significantly predictive of sexual recidivism. It appeared that the predictive power of Mental disorder was primarily due to two of the risk factors included in this domain, Psychopathic personality disorder and Sexual deviance. For Manageability, the significant predictability of this domain also appeared to be driven by two risk factors, Problems with planning and Problems with supervision.

The RSVP total scores were found to have good predictive validity; total scores were significantly predictive of sexual recidivism across the different types of analyses (AUC and Cox regression). There appeared to be no notable difference between the predictive accuracy for total scores of the RSVP, SVR-20, Static-99R, Static-2002R and

the SORAG; each total score significantly predicted sexual recidivism and appeared to perform similarly. Further, there were no significant differences between AUCs for the risk assessment instruments' total and Five Group scores, or risk categories. With regards incremental value, only Static-2002R Five Group scores appeared to significantly add to the predictive accuracy of RSVP Five Group total scores.

The RSVP Summary Risk rating Case Prioritization was associated with good predictive validity for sexual recidivism across the different type of analyses (AUC, Chi Square, odd ratios, Kaplan Meirer Survival Analysis and Cox regression). The predictive accuracy of Case Prioritization appeared essentially equivalent to the predictive accuracy associated with actuarial instruments' risk categories (based on AUC analyses). With respect to odds ratios, the general pattern for all of the sexual violence risk assessment instruments was that offenders in the highest risk category and/or with the highest total scores had the greatest likelihood of sexually recidivism. The only exception was the SORAG where a more moderate bin (risk category) was associated with significant risk for sexual recidivism as opposed to the participants in the highest bin (risk category). With regards to incremental value, none of the other instrument's risk categories appeared to significantly add to the predictive accuracy of Case Prioritization. In terms of the other RSVP Summary Risk Ratings, Immediate Action was significantly correlated with sexual recidivism but Serious Harm was not.

Limitations

Despite the overall positive results of this study, there were a number of limitations that may have impacted the results and the conclusions that can be drawn. This study was conducted using a retrospective file-based research design. The reliance on file-based information may have led to bias in terms of what sort of information was included (or omitted). For example, dynamic risk factors were coded based on someone else's written description of the participant as opposed to being coded based on both file and interview information. Presumably, interview information would be more in depth and accurate than relying solely on a written description of participants. Had interview information been available, various risk factors may have been coded differently, which may have impacted the results of this study. One positive aspect of this study though

was that there were recent sex offender treatment notes for each participant that commented on the dynamic risk factors included in the RSVP.

Another disadvantage to the type of research design used in my study is that I was not able to track and/or consider changes in dynamic risk factors over time. One of the cornerstones of the structured professional judgement approach to violence assessment is that risk is dynamic and needs to be reassessed periodically in order for the risk assessment to remain accurate. When changes in dynamic risk factors are not evaluated over the follow-up period, as was the case in my study, this may lead to an underestimation of the predictive validity of structured professional judgement ratings (Rettenberger & Hucker, 2011). Therefore, it is possible that the RSVP's predictive validity, particularly with regards to individual risk factors, may have been limited by the risk factors only having been coded once at the beginning of the follow-up period.

As previously noted, the majority of the offenders in my sample did not have a previous sexual violence conviction prior to the index offence and the sample was generally rated as low to moderate risk for sexual recidivism. Given these two factors, it is possible that the amount of file information that was available was limited relative to the amount of information that may have been available for a higher risk sample. A higher risk sample would likely include a significant number of participants who had a previous sexual conviction and therefore, the file would likely include more information on the participant. Again, with additional information, the accuracy of the risk factor ratings would likely increase, which in turn would likely impact the predictive validity of the sexual violence risk assessment instruments. Although the predictive validity of the RSVP may have been improved with greater file information, the file information for this study's participants was generally sufficient to code both the RSVP and the other sexual violence risk assessment instruments. Another issue with a relatively lower risk sample is that there was a restriction in the range of ratings for individual risk factors and the numerical scores based on them, as well as in the range of Summary Risk ratings. This range restriction may have resulted in a lower-bound estimate of the interrater reliability, concurrent validity, and predictive validity of decisions made using the RSVP (as well as those made using the other risk assessment instruments).

Previous research has found that the predictive validity of sexual violence risk

assessments may vary across different groups of offenders (e.g., Bartosh et al., 2003; Rettenberger & Eher, 2007). The sample may be divided into groups such as offenders with different types of victims (i.e., children versus adults) or different types of offenders (i.e., contact versus non-contact offences). For my study, although I have the data regarding the breakdown of my sample for different types of victims, due to a relatively small sample size, I did not examine whether the predictive validity of the RSVP varied by sex offender type. It would have been helpful to have a larger sample size so as to examine whether the predictive validity of the RSVP varied across offender type. Previous research (Rettenberger et al., 2010) suggests that standardized risk assessment instruments tend to reflect relatively better predictive accuracy for extra-familial child molesters compared to other sex offender subgroups. This is consistent with research done by Rettenberger, Boer, & Reinhard (2011) who looked at the predictive accuracy of the SVR-20 across different subgroups of sex offenders (*N* = 493). These authors found that in general, the SVR-20 was a better predictor of sexual recidivism for child molesters than rapists. Given the significant overlap between the SVR-20 and RSVP, one may speculate that the RSVP may also be a relatively stronger predictor of sexual recidivism for child molesters in comparison to other subgroups of sex offenders. This is an area that would be beneficial to explore in terms of future research.

The results of this study may also have been limited by the experience of the raters. Although the raters were trained and included one of the authors of the RSVP, not all of the raters had real world clinical experience doing risk assessments. As already discussed earlier, the level of raters' expertise tends to be positively correlated with the interrater reliability of risk ratings (Hanson et al., 2014). Further, a positive association has been found between the interrater reliability and predictive accuracy of risk items for sexual recidivism (Smid et al., 2014). The interrater reliability for this study was excellent for all of the risk assessment instruments in terms of total scores and risk categories, although the interrater reliability was variable across the individual risk factors ranging from fair to excellent. It is possible that the interrater reliability of the individual risk factors may have improved had the raters had a wider breadth of experience doing sexual violence risk assessments. In turn, the predictive accuracy of the individual risk factors, total scores and risk categories may also have increased for the various risk

assessment instruments used in this study.

Implications for Theory

There were a number of individual RSVP risk factors that were coded as present for the majority of the sample. Yet, only one of the most frequent risk factors (Presence Problems with planning) was significantly correlated with sexual recidivism. None of the most frequent risk factors for RSVP Presence and Relevance were significantly predictive of sexual recidivism based on AUCs. Therefore, when conceptualizing risk factors, it is important to separate the frequency of a risk factor from the predictive validity of it. A risk factor may be commonly present among sex offenders yet have little predictive value regarding sexual recidivism. Conversely, a risk factor may be relatively uncommon but be strongly associated with sexual recidivism. With regards to Problems with planning, it is interesting that this risk factor was found to be significantly correlated with sexual recidivism due to its seemingly dynamic nature. This finding suggests that difficulties with planning may tend to be chronic over time and perhaps more characterological in nature than may appear on the surface. It is also noteworthy that whether a sexual offender has problems with planning does not appear to be typically included in risk assessment instruments, particularly actuarial measures.

The majority of the offenders in this sample were deemed to be at low to moderate risk for sexual recidivism and low risk for seriously harming others and imminent sexual violence. Within the literature, it appears that the characteristics of the overall sample, for example, level of risk and history of sexual offending, may have an impact on the results, particularly with regards to predicting sexual recidivism (Darjee et al., 2014; Harris & Hanson, 2004). This raises the question of whether risk factors differ in terms of predictive power with different levels of risk. In other words, would the results of my study have been the same with a higher risk sample, specifically in relation to RSVP individual risk factors? It would seem reasonable that RSVP risk factors would more strongly predict sexual recidivism in a higher risk sample in part because there would likely be more information or evidence on which to base the ratings. High risk sex offenders are more likely to have a well-documented history of sexual offending. Presumably, the more evidence there is to support the absence or presence of risk

factors, the greater the accuracy of the ratings. This notion is consistent research (Guy & Douglas, 2006) that has found that accuracy of risk ratings tends to increase with the availability of additional information (i.e., interview and file data versus file data only).

It appears that the theory of what constitutes a good violence risk assessment is moving beyond simply predicting risk toward including risk formulation and management recommendations (Hart & Logan, 2011). The RSVP clearly fits nicely within this shift given that it provides assessors with a framework for estimating the level of intervention necessary to reduce an offender's risk for sexual recidivism. In addition, it includes steps for identifying likely scenarios for future sexual violence, including factors such as the nature, severity, and imminence of sexual violence, as well as treatment or management strategies (Hart et al., 2003). The expansion of what constitutes a good violence risk assessment challenges the argument that actuarial risk assessment instruments are superior to structured professional judgement guidelines. Actuarial instruments, such as the Static-99R, tend to focus on reporting offenders' scores in terms of percentiles and risk ratios, and do not provide risk formulation or management recommendations. Further, in terms of psychometric properties, the results of this study generally do not support the position that actuarial instruments perform better than structured professional judgement guidelines.

Transparency is important to the risk assessment process for a number of reasons, one being that it should be clear how risk factors are to be coded. The results of this study suggest that the coding guidelines for the RSVP are at least as clear as the coding guidelines for the actuarial instruments in that the average interrater reliability for individual risk factors was fairly consistent across instruments. It has been stated that "an empirically based actuarial approach can be considered a transparent approach to sex offender risk assessment" (Tully et al., 2013, p. 288). However, despite the focus on static or historical risk factors, there remains some ambiguity and subjectivity in the actuarial approach. This is evident by the periodic revision of coding manuals, which provide greater clarification as to how a risk factor is to be coded. For example, the original Static-99 coding manual was revised in 2003 (Harris et al.) in response to the need for greater coding clarification. In addition, if the coding of actuarial instruments was without any ambiguity, then one would expect that the interrater reliability for individual items would consistently be perfect across different raters, which is not

supported by either this study or previous research (Quesada et al., 2014). In terms of conceptualizing transparency within the risk assessment process, it would be helpful for the literature to acknowledge that both the actuarial and structured professional judgement approach provide relatively clear guidelines for coding risk factors. Yet despite this, both approaches have an element of subjectivity when coding risk factors.

Implications for Policy and Practice

Based on the results of this study, the RSVP appears to perform fairly well in terms of interrater reliability, concurrent validity and predictive validity, and to be psychometrically comparable with the other, widely used, sexual violence risk assessment instruments. These findings suggest that the RSVP is appropriate for routine use in sexual violence risk assessments. As noted earlier, one of the main advantages of the RSVP is that it helps to identify key issues related to risk formulation and treatment/management recommendations. This is an important advantage as the purpose for violence risk assessments often includes the need for an understanding of offenders' treatment and/or management needs so as to reduce their risk for reoffending. The RSVP is also versatile in that it can be used at various stages of the legal process (e.g., sentencing or parole) or in different settings (e.g., inpatient or outpatient settings). One consideration though in terms of using the RSVP is that a good number of the risk factors are dynamic in nature. The presentation of dynamic risk factors can change quickly and therefore; the offender's level of risk can also change quickly. The implications of this are that the accuracy of the assessment is contingent on how recently the RSVP was coded, and speaks to the need for periodic reassessment of the offender so as to monitor changes in risk level.

The RSVP interrater reliability in this study was fairly consistent with previous research (Watt et al., 2006), that also found generally good to excellent interrater reliability. However, other research (Sutherland et al., 2012; Wilson, 2013) reported more variable RSVP interrater reliability and indicated that the level of training of the raters was related to interrater reliability. Further, there is a positive relationship between interrater reliability and predictive accuracy for sexual violence risk factors (Smid et al., 2014). These findings highlight the importance of assessors being well trained on the

RSVP and other sexual violence risk assessment instruments.

The present study considered both RSVP total scores and Summary Risk ratings with regards to predictive accuracy, which is consistent with previous research in this area (Dargee et al., 2016; Wilson, 2013). The results of the current study suggest that RSVP total scores and Case Prioritization had similar predictive accuracy. These findings may lead someone to assume that RSVP total scores can be relied on in terms of implying a certain level of risk. While it is true that in general, the more risk factors present, the higher the total score and also the greater the risk level for an offender. However, focusing on totals scores negates the reality that an offender may have only a few risk factors but still be at high risk for sexual recidivism. Therefore, it is important that assessors using the RSVP and other structured professional guidelines not rely on total scores to estimate risk.

The relationship between age and sexual offending appears to be unclear (Lussier & Cale, 2013) although some research (e.g., Hanson, 2001) suggests that the effects of age are dependent on the type of sex offender. Although actuarial instruments typically take type of sex offence and age into consideration as main effects, they do not take the interaction between these factors into account. The RSVP and other structured professional judgement guidelines are in a better position to consider whether there appears to be an interaction between two such factors, and whether there are mitigating factors, such as illness, impacting an offender's level of risk. Further, a structured professional judgement approach is able to consider other individual characteristics such as race, and disability, which may influence the extent to which risk factors contribute to future sexual violence.

It has been recommended that both actuarial and structured professional judgement instruments are utilized as part of sexual violence risk assessments (Rettenberger & Hucker, 2011). In terms of determining the appropriate type of instrument(s) to be used in a violence risk assessment, it is important to take the referral question(s) into consideration. For example, a referral may simply request an assessment determine the likelihood that an offender will sexually reoffend. In this case, the results of this study suggest that it may be redundant and unnecessary to include both an RSVP and actuarial risk assessment instrument in a sexual violence risk

assessment given that both types of instruments performed equally well overall and showed good concurrent validity. In other words, the predictive accuracy and risk ratings were generally consistent between the RSVP and the actuarial instruments. However, different risk assessment approaches may not always have high agreement on the same individual (Jung, Pham, & Ennis, 2013). If multiple risk assessment instruments are used in a sexual violence risk assessment, it would be helpful to include the offender's level of risk for each instrument and an explanation for how the final risk rating was obtained. For example, if discrepancies exist between the risk ratings, how did the evaluator resolve these differences and determine the offender's level of risk for future sexual violence? In the case where the referral for the sexual violence risk assessment is requesting an estimate of risk to reoffend and also treatment and management recommendations, then it may be appropriate to use both actuarial and structured professional judgement instruments. However, given that structured professional judgement instruments, such as the RSVP, are able to address both risk and treatment/management strategies, it appears that using both an actuarial and structured professional judgement approach is unnecessary.

Implications for Future Research

For this study, the process of consensus ratings highlighted the fact that raters would often code different information for the same risk factor. The result being that the consensus rating would be based on more complete information than the individual ratings. It seems likely that the consensus process increased both the predictive accuracy and interrater reliability results in this study. Currently, it appears that consensus ratings are seldom used in research perhaps because of the additional work associated with having two raters code each participant and come to a consensus regarding their ratings. Given the potential benefits though, it would be advantageous for future research to include consensus ratings when possible.

Ideally, future research on the RSVP would include a prospective, longitudinal design based on both clinical interviews and comprehensive file information, as was done by Darjee et al. (2016). As illustrated by Darjee et al. (2016), this type of research allows for the examination of whether structured professional judgements are helpful in

predicting the imminence, severity, nature, and scenarios for future sexual violence. Such research could also consider the role of risk management, formulations and interventions in the process of risk assessment and recidivism. For example, it would be helpful if future research explored to what extent risk management recommendations are implemented and whether the recommendations led to a reduction in future sexual violence. Finally, this type of study design would allow for the periodical assessment of participants to evaluate change in dynamic risk factors.

As discussed earlier, multiple violence risk instruments are often integrated for the purpose of violence risk assessments (Blais & Forth, 2014). Yet it is still relatively ambiguous as to how final risk ratings are made when there are discrepancies across the instruments in terms of risk for sexual recidivism. Research has begun to emerge in this area (Storey, Watt, Jackson, & Hart, 2012), although more work is needed. Such research could help to inform a) whether it is appropriate to use multiple instruments in an assessment and b) if it is appropriate, how should final risk ratings be determined when there are discrepancies among the instruments? Resolving these issues would advance violence risk assessment in part by adding to the transparency of the process.

REFERENCES

Anderson, D., & Hanson, R. K. (2010). Static-99: An actuarial tool to assess risk of sexual and violent recidivism among sexual offenders. In R. K. Otto, & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 251-267). New York, NY, US: Routledge/Taylor & Francis Group.

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, *87*(1), 84-94. doi:10.1207/s15327752jpa8701_07

Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Even highly correlated measures can add incrementally to predicting recidivism among sex offenders. *Assessment*, *19*(4), 442-461. doi:10.1177/1073191112458312

Barbaree, H. E. (2005). Psychopathy, Treatment Behavior, and Recidivism: An Extended Follow-Up of Seto and Barbaree. *Journal of Interpersonal Violence*, *20*(9), 1115-1131. doi:10.1177/0886260505278262

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. *The International Journal of Forensic Mental Health*, *7*(1), 47-64. doi:10.1080/14999013.2008.9914403

Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology*, *47*(4), 422-438. doi:10.1177/0306624X03253850

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: Journal of Research and Treatment*, *22*(2), 234-251. doi:10.1177/1079063210369014

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*(2), 85-106. doi:10.1080/10683160701483104

Blais, J., & Forth, A. E. (2014). Prosecution-retained versus court-appointed experts: Comparing and contrasting risk assessment reports in preventative detention hearings. *Law and Human Behavior*, *38*(6), 531-543. doi:10.1037/lhb0000082

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for Sexual Violence Risk-20: Professional guidelines for assessing risk of sexual violence.* Vancouver, British Columbia: British Columbia Institute on Family Violence and Mental Health, Law, and Policy Institute, Simon Fraser University.

Breiding, M. J. (2015). Prevalence and characteristics of sexual violence, stalking, and intimate partner violence victimization - National Intimate Partner and Sexual Violence Survey, United States, 2011. *American Journal of Public Health*, *105*(4), e11-e12. doi:10.2105/AJPH.2015.302634

Brouillette-Alarie, S., & Proulx, J. (2013). Predictive validity of the Static-99R and its dimensions. *Journal of Sexual Aggression*, *19*(3), 311-328. doi:10.1080/13552600.2012.747630

Campbell, R., Dworkin, E., & Cabral, G. (2009). An ecological model of the impact of sexual assault on women's mental health. *Trauma, Violence, & Abuse*, *10*(3), 225-246. doi:10.1177/1524838009334456

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, *36*(6), 567-590. doi:10.1177/0093854809333610

Cooke, D. J., Michie, C., Hart, S. D., & Hare, R. D. (1999). Evaluating the Screening Version of the Hare Psychopathy Checklist—Revised (PCL:SV): An item response theory analysis. *Psychological Assessment*, *11*(1), 3-13. doi:10.1037/1040-3590.11.1.3

Conroy, M. A., & Witt, P. H. (2013). Evaluation and management of sexual offenders. In R. K. Otto & I.B. Weiner (Eds.), *Handbook of psychology: Forensic psychology (Vol. 11)* (2nd ed.) (pp. 332-356). Hoboken, NJ, US: John Wiley & Sons Inc.

Coxell, A. W., & King, M. B. (2010). Adult male rape and sexual assault: Prevalence, re-victimisation and the tonic immobility response. *Sexual and Relationship Therapy*, *25*(4), 372-379.

Craig, L. A., Beech, A., & Browne, K. D. (2006). Cross-Validation of the Risk Matrix 2000 Sexual and Violent Scales. *Journal of Interpersonal Violence*, *21*(5), 612-633. doi:10.1177/0886260506286876

Craig, L. A., Beech, A. R., & Harkins, L. (2009). The predictive accuracy of risk factors and frameworks. In A. R. Beech, L. A. Craig, & K. D. Browne (Eds.), *Assessment and treatment of sex offenders: A handbook* (pp. 53-74). New York, NY, US: John Wiley & Sons Ltd.

Darjee, R., & Russell, K. (2012). What clinicians need to know before assessing risk in sexual offenders. *Advances in Psychiatric Treatment,18,* 467-478. doi:10.1192/apt.bp.110.008094

Darjee, R., Russel, K., Forrest, L., Milton, L., Savoie, V., Baron, E., … Stobie, S. (2016, January). *Risk for Sexual Violence Protocol: A real world study of the reliability, validity and utility of a structured professional judgment instrument in the assessment and management of sexual offenders in South East Scotland.* Retrieved from http://www.rmascotland.gov.uk/news-and-information/latest-news/risk-sexual-violence-protocol-rsvp-publication-real-world-study-reliability-validity-and-utility-rsvp/

Davidson, J. T., & Chesney-Lind, M. (2009). Discounting women: Context matters in risk and need assessment. *Critical Criminology, 17,* 221-245. doi:10.1007/s10612-009-9084-x

de Vogel, V., de Ruiter, C., van Beek, D., & Mead, G. (2004). Predictive validity of the SVR-20 and Static-99 in a Dutch sample of treated sex offenders. *Law and Human Behavior*, *28*(3), 235-251. doi:10.1023/B:LAHU.0000029137.41974.eb

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.

Dempster, R. J. & Hart, S. D. (2002). The relative utility of fixed and variable risk factors in discriminating sexual recidivists and nonrecidivists. *Sexual Abuse: A Journal of Research and Treatment, 14*, 121-138. doi:10.1023/A:1014668130835

Doren, D. M. (2002). *Evaluating sex offenders: A manual for civil commitments and beyond*. Thousand Oaks, CA, US: Sage Publications, Inc.

Ducro, C., & Pham, T. (2006). Evaluation of the SORAG and the Static-99 on Belgian Sex Offenders Committed to a Forensic Facility. *Sexual Abuse: Journal of Research and Treatment*, *18*(1), 15-26. doi:10.1177/107906320601800102

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: Journal of Research and Treatment*, *24*(4), 350-377. doi:10.1177/1079063211429470

Fabian, J. M. (2006). A literature review of the utility of selected violence and sexual violence risk assessment instruments. *Journal of Psychiatry & Law*, *34*(3), 307-350. doi: 10.1177/009318530603400304

Fazel, S., Sjöstedt, G., Långström, N., & Grann, M. (2006). Risk Factors for Criminal Recidivism in Older Sexual Offenders. *Sexual Abuse: Journal of Research and Treatment*, *18*(2), 159-167. doi:10.1177/107906320601800204

Finkelhor, D. (1994). The international epidemiology on child sexual abuse. *Child Abuse & Neglect, 18*(5), 409-417. doi:10.1016/0145-2134(94)90026-4

Finkelhor, D., Araji, S., Baron, L., Browne, A., Peters, S. D., & Wyatt, G. E. (1986). *A sourcebook on child sexual abuse*. Thousand Oaks, CA, US: Sage Publications, Inc.

Finkelhor, D, & Russell, D. (1984). Women as perpetrators: Review of the evidence. In D. Finkelhor (Ed.), *Child sexual abuse: New theory and research* (pp. 171-187). New York: The Free Press.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*(2), 293-323. doi:10.1037/1076-8971.2.2.293

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*(1), 19-30. doi:10.1037/1040-3590.12.1.19

Guy, L. S. (2010). Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey. *Dissertation Abstracts International*, *71*, 2671.

Guy, L. S., & Douglas, K. S. (2006). Examining the utility of the PCL:SV as a screening measure using competing factor models of psychopathy. *Psychological Assessment*, *18*(2), 225-230. doi:10.1037/1040-3590.18.2.225

Hamby, S. L., & Koss, M. P. (2003). Shades of gray: A qualitative study of terms used in the measurement of sexual victimization. *Psychology of Women Quarterly*, *27*(3), 243-255. doi:10.1111/1471-6402.00104

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29-36. doi: 10.1148/radiology.143.1.7063747

Hannah-Moffat, K. (2013). Actuarial sentencing: An 'unsettled' proposition. *Justice Quarterly, 30*(2), 270-296. doi:10.1080/07418825.2012.682603

Hanson, R. K. (2001). *Age and sexual recidivism: A comparison of rapists and child molesters.* User Report No. 2001-01. Ottawa: Department of the Solicitor General of Canada.

Hanson, R. K. (2002). Recidivism and age: Follow-up data from 4,673 sexual offenders. *Journal of Interpersonal Violence*, *17*(10), 1046-1062. doi:10.1177/088626002236659

Hanson, R. K., Babchishin, K. M., Helmus, L., & Thornton, D. (2013). Quantifying the relative risk of sex offenders: Risk ratios for Static-99R. *Sexual Abuse: Journal of Research and Treatment*, *25*(5), 482-515. doi:10.1177/1079063212469060

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, *66*(2), 348-362. doi:10.1037/0022-006X.66.2.348

Hanson, R. K., Harris, A. R., Helmus, L., & Thornton, D. (2014). High-risk sex offenders may not be high risk forever. *Journal of Interpersonal Violence*, *29*(15), 2792-2813. doi:10.1177/0886260514526062

Hanson, R. K., Harris, A. R., Scott, T., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project.* Corrections User Report No 2007–05. Ottawa, Canada: Public Safety Canada.

Hanson, R. K., Lloyd, C. D., Helmus, L., & Thornton, D. (2012). Developing non-arbitrary metrics for risk communication: Percentile ranks for the Static-99/R and Static-2002/R sexual offender risk tools. *The International Journal of Forensic Mental Health*, *11*(1), 9-23. doi:10.1080/14999013.2012.667511

Hanson, R. K., Lunetta, A., Phenix, A., Neeley, J., & Epperson, D. (2014). The field validity of Static-99/R sex offender risk assessment tool in California. Journal of Threat Assessment and Management, 1(2), 102-117. doi:10.1037/tam0000014

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), 1-21. doi:10.1037/a0014421

Hanson, R. K., & Thornton, D. M. (1999). *Static-99: Improving actuarial risk assessment for sexual offenders*. Corrections Research User Report, 1999-02. Ottawa, ON, Canada: Department of the Solicitor General of Canada.

Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static-2002.* Corrections Research User Report No. 2003–01. Ottawa, ON, Canada: Department of the Solicitor General of Canada.

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

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

Harris, A. R., & Hanson, R. K. (2004). *Sex offender recidivism: A simple question.* User Report 2004-03. Ottawa, ON, Canada: Public Safety and Emergency Preparedness Canada.

Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. M. (2003). *Static-99 Coding Rules Revised – 2003.* Corrections Research User Report 2003. Ottawa, ON, Canada: Department of the Solicitor General of Canada.

Harris, G. T., Rice, M. E., Quinsey, V. L., & Cormier, C. A. (2015). *Violent offenders: Appraising and managing risk* (3rd ed.). Washington, DC, US: American Psychological Association.

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment*, *15*(3), 413-425. doi:10.1037/1040-3590.15.3.413

Hart, S. D. (2003a). Actuarial risk assessment: Commentary on Berlin et al. *Sexual Abuse: Journal of Research and Treatment, 15*(4)*,* 383-388. doi:10.1177/107906320301500413

Hart, S. D. (2003b, April*). Assessing risk for sexual violence: The Risk for Sexual Violence Protocol (RSVP)*. Paper presented at the Annual Meeting of the International Association of Forensic Mental Health Services, Miami, FL.

Hart, S. D., & Boer, D. P. (2010). Structured professional judgment guidelines for sexual violence risk assessment: The Sexual Violence Risk-20 (SVR-20) and Risk for Sexual Violence Protocol (RSVP). In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 269-294). New York, NY, US: Routledge/Taylor & Francis Group.

Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *Manual for the Hare Psychopathy Checklist: Screening Version (PCL:SV)*. Toronto ON, Canada: Multi-Health Systems.

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

Hart, S. D., Kropp, P. R., Laws, D. R., Klaver, J., Logan, C., & Watt, K. A. (2003). *The Risk for Sexual Violence Protocol (RSVP): Structured professional guidelines for assessing risk for sexual violence.* Vancouver, BC: The Institute Against Family Violence.

Hart, S. D., & Logan, C. (2011). Formulation of violence risk using evidence based assessments: The structured professional judgement approach. In P. Sturney & M. McMurran (Eds.), *Forensic case formulation* (pp. 83-106). Chichester, UK: Wiley-Blackwell. doi: 10.1002/9781119977018.ch4

Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. 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 and Behavior*, *39*(9), 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: Journal of Research and Treatment*, *24*(1), 64-101. doi:10.1177/1079063211409951

Hildebrand, M., de Ruiter, C., & de Vogel, V. (2004). Psychopathy and Sexual Deviance in Treated Rapists: Association With Sexual and Nonsexual Recidivism. *Sexual Abuse: Journal of Research and Treatment*, *16*(1), 1-24. doi:10.1177/107906320401600101

Hill, A., Habermann, N., Klusmann, D., Berner, W., & Briken, P. (2008). Criminal recidivism in sexual homicide perpetrators. *International Journal of Offender Therapy and Comparative Criminology, 52*(1), 5-20. doi: 10.1177/0306624X07307450

Isely, P. J., Busse, W., & Isely, P. (1998). Sexual assault of males in late adolescence: A hidden phenomenon. *Professional School Counseling, 2*(2), 153-160.

Jackson, K., & Healey, J. (2008, July). *Concurrent validity of the RSVP vis-à-vis the SVR-20, Static-99, Static-2002, and SORAG*. Paper presented at the Annual Meeting of the International Association of Forensic Mental Health Services, Vienna, Austria.

Janka, C., Gallasch-Nemitz, F., Biedermann, J., & Dahle, K. (2012). The significance of offending behavior for predicting sexual recidivism among sex offenders of various age groups. *International Journal of Law and Psychiatry, 35*(3), 159-164. doi:10.1016/j.ijlp.2012.02.002

Judge, J., Quayle, E., O'Rourke, S., Russell, K., & Darjee, R. (2014). Referrers' views of structured professional judgement risk assessment of sexual offenders: A qualitative study. *Journal of Sexual Aggression, 20*(1), 94-109. doi:10.1080/13552600.2013.767948

Khiroya, R., Weaver, T., & Maden, T. (2009). Use and perceived utility of structured violence risk assessments in English medium secure forensic units. *Psychiatric Bulletin, 33*(4), 129-132. doi:10.1192/pb.bp.108.019810

Klaver, J., Watt, K., Kropp, P. R., & Hart, S. D. (2002, August). *Actuarial assessment of risk for sexual violence.* Paper presented at the annual meeting of the American Psychological Association, Chicago, IL.

Kropp, P. R. (2001, April). *The Risk for Sexual Violence Protocol (RSVP).* Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Vancouver, BC.

Långström, N. (2004). Accuracy of actuarial procedures for assessment of sexual offender recidivism risk may vary across ethnicity. *Sexual Abuse: Journal of Research and Treatment, 16*(2), 107-120. doi:10.1177/107906320401600202

Lievore, D. (2003). *Non-reporting and Hidden Recording of Sexual Assault: An International Literature Review*. Commonwealth Office of the Status of Women, Canberra.

Looman, J. (2006). Comparison of two risk assessment instruments for sexual offenders. *Sexual Abuse: Journal of Research and Treatment, 18*(2), 193-206. doi:10.1177/107906320601800206

Lund, C., Hofvander, B., Forsman, A., Anckarsäter, H., & Nilsson, T. (2013). Violent criminal recidivism in mentally disordered offenders: A follow up study of 13–20 years through different sanctions. *International Journal of Law and Psychiatry, 36*(3-4), 250-257. doi:10.1016/j.ijlp.2013.04.015

Lussier, P., & Blokland, A. (2014). The adolescence-adulthood transition and robins' continuity paradox: Criminal career patterns of juvenile and adult sex offenders in a prospective longitudinal birth cohort study. *Journal of Criminal Justice, 42*(2), 153-163. doi:10.1016/j.jcrimjus.2013.07.004

Lussier, P., & Cale, J. (2013). Beyond sexual recidivism: A review of the sexual criminal career parameters of adult sex offenders. *Aggression and Violent Behavior*, *18*(5), 445-457. doi:10.1016/j.avb.2013.06.005

Lussier, P., Van Den Berg, C., Bijleveld, C., & Hendriks, J. (2012). A developmental taxonomy of juvenile sex offenders for theory, research, and prevention: The adolescent-limited and the high-rate slow desister. *Criminal Justice and Behavior*, *39*(12), 1559-1581. doi:10.1177/0093854812455739

Mathews, F. (1987). *Adolescent sex offenders: A needs study*. Ministry of Community and Social Services. Toronto, ON: Central Toronto Youth Services.

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: Journal of Research and Treatment*, *24*(5), 431-458. doi:10.1177/1079063211432475

Monahan, J. M., Steadman, H. J., Silver, E., Appelbaum, P. S., Clark Robbins, P., Mulvey, E. P.,… Banks, S. (2001). *Rethinking risk assessment: The MacArthur Study of mental disorder and violence.* New York, NY: Oxford University Press: 2001.

Motiuk, L. L., & Belcourt, R. (1996). Profiling the Canadian federal sex offender population. *Forum on Corrections Research, 8,* 3-7. Retrieved from http://www.csc-scc.gc.ca/research/forum/e082/082a_e.pdf

Muñoz-Rivas, M. J., Graña, J. L., O'Leary, K. D., & González, M. P. (2009). Prevalence and predictors of sexual aggression in dating relationships of adolescents and young adults. *Psicothema, 21*(2), 234-240.

Nisbet, I. A., Wilson, P. H., & Smallbone, S. W. (2004). A prospective longitudinal study of sexual recidivism among adolescent sex offenders. *Sexual Abuse: Journal of Research and Treatment*, *16*(3), 223-234. doi:10.1177/107906320401600304

O'Brien, R. M. (2007). A caution regarding rules of thumb for Variance Inflation Factors. *Quality & Quantity, 41*, 673-690. doi:10.1007/s11135-006-9018-6

Olver, M. E., Wong, S. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale-Sexual Offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment*, *19*(3), 318-329. doi:10.1037/1040-3590.19.3.318

Parent, G., Guay, J., & Knight, R. A. (2011). An assessment of long-term risk of recidivism by adult sex offenders: One size doesn't fit all. *Criminal Justice and Behavior*, *38*(2), 188-209. doi:10.1177/0093854810388238

Pedersen, L., Rasmussen, K., & Elsass, P. (2010). Risk assessment: The value of structured professional judgments. *The International Journal of Forensic Mental Health*, *9*(2), 74-81. doi:10.1080/14999013.2010.499556

Phenix, A., Helmus, L., & Hanson, R. K. (2015). *Static-99R & Static-2002R evaluator's workbook*. Retrieved from http://www.static99.org/pdfdocs/Static-99RandStatic-2002R_EvaluatorsWorkbook-Jan2015.pdf

Prentky, R. A., Lee, A. F. S., Knight, R. A., & Cerces, D. (1997). Recidivism rates in child molesters and rapists: A methodological analysis. *Law and Human Behavior, 21*, 635–659.

Prescott, D. S. (2004). Book review: The Risk for Sexual Violence Protocol (RSVP). *ATSA Forum Newsletter, Fall 2004*. Retrieved from http://www.davidprescott.net/pub_17.shtml.

Quesada, S. P., Calkins, C., & Jeglic, E. L. (2014). An examination of the interrater reliability between practitioners and researchers on the Static-99. *International Journal of Offender Therapy and Comparative Criminology*, *58*(11), 1364-1375. doi:10.1177/0306624X13495504

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, US: American Psychological Association. doi:10.1037/11367-000

Ramírez, M. P., Illescas, S. R., García, M. M., Forero, C. G., & Pueyo, A. A. (2008). Predicción de riesgo de reincidencia en agresores sexuales (Recidivism risk assessment in sex offenders). *Psicothema, 20*(2), 205-210.

Randall, M., & Haskell, L. (1995). Sexual violence in women's lives: Findings from the Women's Safety Project, a community-based survey. *Violence Against Women, 1*(1), 6-31.

Rettenberger, M., Boer, D. P., & Eher, R. (2011). The predictive accuracy of risk factors in the Sexual Violence Risk–20 (SVR-20*). Criminal Justice and Behavior, 38*(10), 1009-1027. doi:10.1177/0093854811416908

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. (2011). Structured professional guidelines: International applications. 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. 85-110). Chichester, UK: Wiley-Blackwell. doi:10.1002/9781119990420.ch5

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 and Comparative Criminology*, *54*(2), 169-186. doi:10.1177/0306624X08328755

Rice, A. K., Boccaccini, M. T., Harris, P. B., & Hawes, S. W. (2014). Does field reliability for Static-99 scores decrease as scores increase? *Psychological Assessment*, *26*(4), 1085-1094. doi:10.1037/pas0000009

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's *d*, and *r*. *Law and Human Behavior*, *29*(5), 615-620. doi:10.1007/s10979-005-6832-7

Sackett, D. L., Rosenberg. W. M., Muir Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn`t. *British Medical Journal, 312,* 71-72. doi:http://dx.doi.org/10.1136/bmj.312.7023.71

Sansone, R., & Sansone, L. (2011). Sexual behaviour in borderline personality: A review. *Innovations in Clinical Neuroscience, 8*(2), 14-18.

Seto, M. C., & Lalumière, M. L. (2010). What is so special about male adolescent sexual offending? A review and test of explanations through meta-analysis. *Psychological Bulletin*, *136*(4), 526-575. doi:10.1037/a0019700

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*(3), 499-513. doi:10.1016/j.cpr.2010.11.009

Sinha, M. (2013). Measuring violence against women: Statistical trends. *Juristat, 85-002-X.* Retrieved from http://www.statcan.gc.ca/pub/85-002-x/2013001/article/11766/11766-1-eng.htm

Sipe, R., Jensen, E. L., & Everett, R. S. (1998). Adolescent sexual offenders grown up: Recidivism in young adulthood. *Criminal Justice and Behavior*, *25*(1), 109-124. doi:10.1177/0093854898025001007

Sjöstedt, G., & Långström, N. (2001). Actuarial assessment of sex offender recidivism risk: A cross-validation of the RRASOR and the Static-99 in Sweden. *Law and Human Behavior*, *25*(6), 629-645. doi:10.1023/A:1012758307983

Sjöstedt, G., & Långström, N. (2002). Assessment of risk for criminal recidivism among rapists: A comparison of four different measures. *Psychology, Crime & Law*, *8*(1), 25-40. doi:10.1080/10683160208401807

Slotboom, A., Hendriks, J., & Verbruggen, J. (2011). Contrasting adolescent female and male sexual aggression: A self-report study on prevalence and predictors of sexual aggression. *Journal of Sexual Aggression*, *17*(1), 15-33. doi:10.1080/13552600.2010.544413

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: Journal of Research and Treatment*, *25*(3), 302-316. doi:10.1177/1079063212472937

Smid, W. J., Kamphuis, J. H., Wever, E. C., & Van Beek, D. J. (2014). A comparison of the predictive properties of nine sex offender risk assessment instruments. *Psychological Assessment*, *26*(3), 691-703. doi:10.1037/a0036616

Song, L., Lieb, R. & Donnelly, S. (1993). *Female sex offenders in Washington State*. Retrieved from http://www.wsipp.wa.gov/ReportFile/1154

Sorenson, S. B., Stein, J. A., Siegel, J. M., Golding, J. M, & Burnam, M. A. (1988). The prevalence of adult sexual assault: The Los Angeles Epidemiologic Catchment Area Project. *American Journal of Epidemiology, 126*(6), 1154-1164

Spiranovic, C. (2012). *The Static-99 and Static-99R norms project: Developing norms based on Western Australian sex offenders.* Perth: Department of Corrective Services (Western Australia). Retrieved from http://www.law.uwa.edu.au/__data/assets/pdf_file/0007/2253328/FinalReportForRelease141212_Static99Norms_141212.pdf

Sreenivasan, S., Weinberger, L. E., Frances, A., & Cusworth-Walker, S. (2010). Alice in actuarial-land: Through the looking glass of changing Static-99 norms*. Journal of the American Academy of Psychiatry and the Law, 38*(3), 400-406.

Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U., & Nedopil, N. (2005). Risk assessment and prediction of violent and sexual recidivism in sex offenders: Long-term predictive validity of four risk assessment instruments. *Journal of Forensic Psychiatry & Psychology, 16(*1), 92-108. doi:10.1080/1478994042000270247

Statistics Canada (2010). Criminal victimization in Canada, 2009. *Juristat, 30(2).* Retrieved from http://www.statcan.gc.ca/pub/85-002-x/2010002/article/11340/tbl/tbl4-eng.htm.

Storey, J. E., Watt, K. A., Jackson, K. J., & Hart, S. D. (2012). Utilization and implications of the Static-99 in practice. *Sexual Abuse: Journal of Research and Treatment*, *24*(3), 289-302.

Sutherland, A. A. (2010). *Sexual violence risk assessment: An investigation into the interrater reliability of the RSVP in Scotland*. Unpublished doctoral dissertation, Section of Psychological Medicine, University of Glasgow, Glasgow, Scotland.

Sutherland, A. A., Johnstone, L., Davidson, K. M., Hart, S. D., Cooke, D. J., Kropp, P. R., & ... Stocks, R. (2012). Sexual violence risk assessment: An investigation of the interrater reliability of professional judgments made using the Risk for Sexual Violence Protocol. *The International Journal of Forensic Mental Health*, *11*(2), 119-133. doi:10.1080/14999013.2012.690020

Thornton, D. (2010). *Scoring guide for the Risk Matrix 2000*. 10/SVC. Unpublished manuscript.

Tjaden, P., & Thoennes, N. (2000). *Extent, nature, and consequences of intimate partner violence: Findings from the National Violence Against Women Survey*. National Institute of Justice and the Centers for Disease Control and Prevention. Washington, DC: Government Printing Office.

Tosh, J., & Carson, K. (In press). Psychiatry's treatment of rape victims: A discursive and intersectional analysis of 'penetration disorder'. *Psychology of Women Section Review.*

Tully, R. J., Chou, S., & Browne, K. D. (2013). A systematic review on the effectiveness of sex offender risk assessment tools in predicting sexual recidivism of adult male sex offenders. *Clinical Psychology Review*, *33*(2), 287-316. doi:10.1016/j.cpr.2012.12.002

Ullman, S. E., & Brecklin, L. R. (2002). Sexual assault history and suicidal behavior in a national sample of women. *Suicide and Life-Threatening Behavior, 32*(2), 117-130. doi:10.1521/suli.32.2.117.24398

Watt, K. A., Hart, S. D., Wilson, C., Guy, L., & Douglas, K. S. (2006, March*). An evaluation of the Risk for Sexual Violence Protocol (RSVP) in high risk offenders: Interrater reliability and concurrent validity.* Paper presented at the Annual Meeting of the American Psychology-Law Society, St. Petersburg, FL.

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.

Zanatta, R. G. (2006). Risk of violent and sexual recidivism: A comparison of dangerous offenders and repetitive sexual offenders. *Dissertation Abstracts International*, *67*, 1723.

Zinger, I. (2004). Actuarial risk assessment and human rights: A commentary. *Canadian Journal of Criminology and Criminal Justice, 46,* 607-620. doi:10.3138/cjccj.46.5.607

Appendices

Appendix A.

RSVP Worksheet

Step 1: Case Information Gather relevant background information	
Name of person being evaluated: Sources reviewed: Synopsis of offence history: Completed by: Date completed:	
Steps 2 & 3 Determine the presence of risk factors prior to and within the last year (Past vs. Recent), as well as their relevance to the development of future management strategies (Future)	
A. Sexual Violence History	**Coding** **Yes, Possibly, No**
1. Chronicity of Sexual Violence	Presence: Past _____ Presence: Recent _____ Presence: Future _____
2. Diversity of Sexual Violence	Presence: Past _____ Presence: Recent _____ Presence: Future _____
3. Escalation of Sexual Violence	Presence: Past _____ Presence: Recent _____ Presence: Future _____
4. Physical Coercion in Sexual Violence	Presence: Past _____ Presence: Recent _____ Presence: Future _____
5. Psychological Coercion in Sexual Violence	Presence: Past _____ Presence: Recent _____ Presence: Future _____

B. Psychological Adjustment	Coding
6. Extreme Minimization or Denial of Sexual Violence	Presence: Past _____ Presence: Recent _____ Presence: Future _____
7. Attitudes that Support or Condone Sexual Violence	Presence: Past _____ Presence: Recent _____ Presence: Future _____
8. Problems with Self-Awareness	Presence: Past _____ Presence: Recent _____ Presence: Future _____
9. Problems with Stress or Coping	Presence: Past _____ Presence: Recent _____ Presence: Future _____
10. Problems Resulting From Child Abuse	Presence: Past _____ Presence: Recent _____ Presence: Future _____
C. Mental Disorder - *For each item in this Domain, specific if Definite or Provisional*	
11. Sexual Deviance	Presence: Past _____ Presence: Recent _____ Presence: Future _____
12. Psychopathic Personality Disorder	Presence: Past _____ Presence: Recent _____ Presence: Future _____
13. Major Mental Illness	Presence: Past _____ Presence: Recent _____ Presence: Future _____
14. Problems with Substance Use	Presence: Past _____ Presence: Recent _____ Presence: Future _____
15. Violent or Suicidal Ideation	Presence: Past _____ Presence: Recent _____ Presence: Future _____

D. Social Adjustment	Coding
16. Problems with Intimate Relationships	Presence: Past _____ Presence: Recent _____ Presence: Future _____
17. Problems with Non-Intimate Relationships	Presence: Past _____ Presence: Recent _____ Presence: Future _____
18. Problems with Employment	Presence: Past _____ Presence: Recent _____ Presence: Future _____
19. Non-Sexual Criminality	Presence: Past _____ Presence: Recent _____ Presence: Future _____
E. Manageability	**Coding**
20. Problems with Planning	Presence: Past _____ Presence: Recent _____ Presence: Future _____
21. Problems with Treatment	Presence: Past _____ Presence: Recent _____ Presence: Future _____
22. Problems with Supervision	Presence: Past _____ Presence: Recent _____ Presence: Future _____
F. Other Considerations	**Coding**
Specify	Presence: Past _____ Presence: Recent _____ Presence: Future _____
Specify	Presence: Past _____ Presence: Recent _____ Presence: Future _____

Formulation of Sexual Violence
Provide an account of the person's sexual violence by identifying primary risk factors and their causal roles. Motivators (factors that increased the perceived likelihood of gains or benefits of violence) Disinhibitors (factors that decreased the perceived likelihood of costs or negative consequences of violence) Destabilizers (factors that impaired, disturbed, or disrupted the process of decision making)

Step 4 – Risk Scenarios Identify and describe the most plausible scenarios of future sexual violence; revise Step 3 as required			
	Scenario #1	**Scenario #2**	**Scenario #3**
Nature What kind of sexual violence is the person likely to commit? Who are the likely victims? What is the likely motivation – that is, what might the person be trying to accomplish?			
Severity What would be the psychological harm to victims? What would be the physical harm to victims? Is there a chance that the sexual violence might escalate to serious or life-threatening violence?			
Imminence How soon might the person engage in sexual violence? Are there any warning signs that might signal that the risk is increasing or imminent?			
Frequency/Duration How often might the sexual violence occur – once, several times, frequently? Is the risk chronic or acute (i.e., time-limited?)			
Likelihood In general, how frequent or common is this type of sexual violence? Based on this person's history, how likely is it that this type of sexual violence will occur?			

Step 5 – Case Management Recommend strategies for managing sexual violence risk; revise Step 3 & 4 as required			
	Scenario #1	**Scenario #2**	**Scenario #3**
Monitoring What is the best way to monitor warning signs that the risk posed by the person may be increasing? What events, occurrences, or circumstances should trigger a re-assessment of risk?			
Treatment What treatment or rehabilitation strategies could be implemented to manage the risks posed by the person? Which deficits in psychological adjustment are high priorities for intervention?			
Supervision What supervision or surveillance strategies could be implemented to manage the risks posed by the person? What restrictions on activity, movement, association, or communication are indicated?			
Victim Safety Planning What steps could be taken to enhance the security of any potential victims? How might the physical security or self-protective skills of potential victims be improved?			
Other considerations What events, occurrences, or circumstances might increase or decrease risk? What else might be done to manage risk?			

Step 6 – Summary Judgements Document judgments regarding overall risk		
Opinion	**Coding**	**Comments**
Summary Risk Rating What level of effort or intervention will be required to prevent further sexual violence? To what extent is this opinion limited in light of information that is unclear, unavailable, or missing?	High/Urgent ☐ Moderate/Elevated ☐ Low/Routine ☐	
Serious Physical Harm What is the risk that the sexual violence will involve or escalate into serious or life-threatening physical harm? To what extent is this opinion limited in light of information that is unclear, unavailable, or missing?	High ☐ Moderate ☐ Low ☐	
Immediate Action Required Does the person pose any imminent risk? What preventative steps were or should be taken immediately?	High ☐ Moderate ☐ Low ☐	
Other Risks Indicated Is there evidence that the person poses other risks, such as general violence, suicide, or self-harm? Should the person be evaluated for other risks?	Yes ☐ Possibly ☐ No ☐	
Case Review When should the case be scheduled for routine review (re-assessment)? What circumstances should trigger a special review (re-assessment)?	**Date for Review:**	

Appendix B.

SVR-20 Worksheet

Assessment of: Name: Age:		
Specify time period for evaluation Recent change:		
Psychosocial Adjustment	**Coding** **No, Possibly, Yes**	**Recent Change** **+, 0, -**
1. Sexual Deviation		
2. Victim of Child Abuse		
3. Psychopathy		
4. Major Mental Illness		
5. Substance Use Problems		
6. Suicidal/homicidal Ideation		
7. Relationship Problems		
8. Employment Problems		
9. Past Non-Sexual Violent Offences		
10. Past nonviolent Offences		
11. Past Supervision Failures		

Sexual Offences	Coding	Recent Change
12. High Density Sex Offences		
13. Multiple Sex Offence Types		
14. Physical Harm to Victim(s) in Sex Offences		
15. Uses Weapons or Threats of Death in Sex Offences		
16. Escalation in Frequency or Severity of Sex Offences		
17. Extreme Minimization or Denial of Sex Offences		
18. Attitudes that Support or Condone Sex Offences		
Future Plans		
19. Lacks Realistic Plans		
20. Negative Attitude Toward Intervention		
Other Considerations		

Summary Risk Rating			
Risk of Sexual Violence	Low ☐	Moderate ☐	High ☐

Evaluation Conducted By:	
Name:	Title:
Signature:	Date:

Appendix C.

Static-99R Worksheet

Question	Risk Factor	Codes		Score
1	Age at release	Aged 18 to 34.9 Aged 35 to 39.9 Aged 40 to 59.9 Aged 60 or older		1 0 -1 -3
2	Ever lived with lover for at least two years?	Yes No		0 1
3	Index non-sexual violence - Any Convictions	No Yes		0 1
4	Prior non-sexual violence - Any Convictions	No Yes		0 1
5	Prior Sex Offences	Charges 0 1,2 3-5 6+	Convictions 0 1 2,3 4+	 0 1 2 3
6	Prior sentencing dates (excluding index)	3 or less 4 or more		0 1
7	Any convictions for non-contact sex offences	No Yes		0 1
8	Any Unrelated Victims	No Yes		0 1
9	Any Stranger Victims	No Yes		0 1
10	Any Male Victims	No Yes		0 1
Total Score - Add up scores from individual risk factors				

Translating Static-99R Scores Into Risk Categories	
Score	**Risk Category**
-3 through 1	Low
2, 3	Low-Moderate
4, 5	Moderate-High
6 plus	High

Appendix D.

Static-2002R Worksheet

Items	Raw Score	Subscore
AGE 1. Age at Release 18 to 34.9 = 2 35 to 39.9 = 1 40 to 59.9 = 0 60 or older = -2		
PERSISTENCE OF SEXUAL OFFENDING 2. Prior Sentencing Occasions for Sexual Offences: No prior sentencing dates for sexual offences = 0 1 = 1 2, 3 = 2 4 or more = 3		
3. Any Juvenile Arrest for a Sexual Offence and Convicted as an Adult for a Separate Sexual Offence: No arrest for a sexual offence prior to age 18 = 0 Arrest prior to age 18 and conviction after age 18 = 1		
4. Rate of Sexual Offending: Less than one sentencing occasion every 15 years = 0 One or more sentencing occasions every 15 years = 1		
Persistence Raw Score (subtotal of Sexual Offending) 0 = 0 1 = 1 2, 3 = 2 4, 5 = 3		
Persistence of Sexual Offending SUBSCORE		
DEVIANT SEXUAL INTERESTS 5. Any Sentencing Occasion For Non-contact Sex Offences: No = 0 Yes = 1		
6. Any Male Victim: No = 0 Yes = 1		
7. Young, Unrelated Victims: Does not have two or more victims < 12, one of them unrelated = 0 Does have two or more victims < 12 years, one must be unrelated = 1		
Deviant Sexual Interest SUBSCORE		
RELATIONSHIP TO VICTIMS 8. Any Unrelated Victim:		

No = 0 Yes = 1		
9. Any Stranger Victim: No = 0 Yes = 1		
Relationship to Victims SUBSCORE		
GENERAL CRIMINALITY 10. Any Prior Involvement with the Criminal Justice System No = 0 Yes = 1		
11. Prior Sentencing Occasions For Anything: 0-2 prior sentencing occasions for anything = 0 3-13 prior sentencing occasions = 1 14 or more prior sentencing occasions = 2=		
12. Any Community Supervision Violation: No = 0 Yes = 1		
13. Years Free Prior to Index Sex Offence: • More than 36 months free prior to committing the sexual offence that resulted in the index conviction AND more than 48 months free prior to index conviction = 0 • Less than 36 months free prior to committing the sexual offence that resulted in the index conviction OR less than 48 months free prior to conviction for index sex offence = 1		
14. Any Prior Non-sexual Violence Sentencing Occasion: No = 0 Yes = 1		
General Criminality raw score (subtotal General Criminality items) 0 = 0 1, 2 = 1 3, 4 = 2 5, 6 = 3		
General Criminality SUBSCORE		
TOTAL -2 to 13		

Translating Static-2002R Scores Into Risk Categories	
Score	**Risk Category**
-2 through 2	Low
3, 4	Low-Moderate
5, 6	Moderate
7, 8	Moderate-High
9 plus	High

Appendix E.

SORAG Worksheet

Risk Factor	Coding
1. Lived with both biological parents to age 16 (except for death of parent).	☐ Yes = -2 ☐ No = +3
2. Elementary school maladjustment (including gr 8)	☐ No problems = -1 ☐ Slight (minor discipline or attendance) or moderate problems = +2 ☐ Severe problems (frequent disruptive behavior and/or attendance or behavior resulting in expulsion or serious suspensions) = +5
3. History of alcohol problems	☐ Parental alcoholism ☐ Teenage alcohol problem ☐ Adult alcohol problem ☐ Alcohol involved in a prior offence ☐ Alcohol involved in the index offence Allot one point for each of the above. 0 points = -1 1 or 2 points = 0 3 points = +1 4 or 5 points = +2
4. Marital status	☐ Ever married (or lived common law in the same home for at least 6 months) = -2 ☐ Never married = +1
5. Criminal history (not including index) score for nonviolent offences (from Cormier-Lang system in Appendix).	☐ Score of 0 = -2 ☐ Score of 1 or 2 = 0 ☐ Score of 3 or over - +3
6. Criminal history (not including index) score for violent offences (from Cormier-Lang system in Appendix)	☐ Score of 0 = -1 ☐ Score of 2 = 0 ☐ Score of 3 or above = +6

7. Number of previous convictions for sexual offences (hands-on sexual offences prior to the index offence)	☐ 0 = -1 ☐ 1 or 2 = +1 ☐ ≥ 3 = +5
8. History of sex offences only against girls under 14 (including index offences; if offender was less than 5 years older than victim, always score +4)	☐ Yes = 0 ☐ No = +4
9. Failure on prior conditional release (includes parole/probation violation or revocation, failure to comply, bail violation, and any new arrest while on conditional release)	☐ No = 0 ☐ Yes = +3
10. Age index offence (at most recent birthday)	☐ ≥ 39 = -5 ☐ 34–38 = -2 ☐ 28–33 = -1 ☐ 27 = 0 ☐ ≤ 26 = +2
11. Meets DSM-III criteria for any personality disorder	☐ No = -2 ☐ Yes = +3
12. Meets DSM-III criteria for schizophrenia	☐ Yes = -3 ☐ No = +1
13. Phallometric test results	☐ All indicate non-deviant sexual preferences = -1 ☐ Any test indicates deviant sexual preferences =+1
14. Psychopathy Checklist score (leave blank, will be filled in later)	☐ ≤ 4 = -5 ☐ 5-9 = -3 ☐ 10-14 = -1 ☐ 15-24 = 0 ☐ 25-34 = +4 ☐ ≥ 35 = + 12
TOTAL	

Appendix F.

The PCL:SV Worksheet

Items	Evidence	Rating			
		0 No	1 Possibly	2 Yes	Omit
1. Superficial					
2. Grandiose					
3. Deceitful					
4. Lacks Remorse					
5. Lacks Empathy					
6. Doesn`t Accept Responsibility					
7. Impulsive					
8. Poor Behaviour Controls					
9. Lacks Goals					
10. Irresponsible					
11. Adolescent Antisocial Behaviour					
12. Adult Antisocial Behaviour					
TOTAL					

Appendix G.

Additional Analyses

Table G1 Interrater Reliability of SVR-20 Ratings, Individual Risk Factors

SVR-20 Items	Score Frequency/ Percentage			ICC_1	ICC_2
	0	1	2		
Sexual deviation	40	29	31	.83	.91
Victim of child abuse	39	10	51	.90	.95
Psychopathy	63	24	13	.76	.86
Major mental illness	50	25	25	.86	.93
Substance use problems	28	12	60	.93	.96
Suicidal/homicidal ideation	54	18	28	.84	.91
Relationship problems	3	16	81	.77	.87
Employment problems	19	39	42	.77	.87
Past nonsexual violent offences	46	23	31	.81	.90
Past nonviolent offences	39	10	51	.81	.89
Past supervision failure	47	4	49	.96	.98
High density	31	7	62	.88	.94
Multiple types	64	25	11	.71	.83
Physical harm	84	9	7	.90	.95
Weapons/threats	93	1	6	.96	.98
Escalation in frequency or severity	54	21	25	.65	.79
Extreme minimization/denial	11	34	55	.68	.81
Attitudes that support or condone	18	43	39	.70	.82
Lacks realistic plans	5	32	63	.64	.78
Negative attitude toward intervention	23	33	44	.74	.85

Note. $N = 100$.

Table G2 ***Interrater Reliability of Static-99R, Static-2000R, and SORAG Total Scores and Risk Categories***

	Total Score		Risk Categories/Bins	
Risk Measure	**ICC_1**	**ICC_2**	**ICC_1**	**ICC_2**
Static-99R	.92	.96	.89	.94
Static-2002R	.92	.96	.88	.93
SORAG	.92	.96	.91	.96

Table G3 ***Frequencey and Interrater Reliability of the Static-99R Ratings, Individual Risk Factors***

Static-99R Items	**Score Frequency**						**ICC_1**	**ICC_2**
	-3	-1	0	1	2	3		
Prior sex offences			64	14	11	11	.73	.84
Prior sentencing dates			71	29			.78	.87
Non-contact sex convictions			76	24			.86	.93
Index non-sex violent conviction			91	9			.50	.67
Prior non-sex violent conviction			78	22			.92	.96
Any unrelated victims			24	76			.69	.82
Any stranger victims			61	39			.77	.87
Any male victims			84	16			.81	.90
Young	7	44	19	30			.88	.94
Ever lived with a partner			73	27			.83	.91

Note. *N* = 100.

Table G4 Frequency and Interrater Reliability of Static-2002R Ratings, Individidual Risk Factors

Static-2002R Items	Score Frequency					ICC_1	ICC_2
	-2	0	1	2	3		
Age at release	7	44	19	30		.95	.98
Prior Sent. Occasions for Sex Offences		68	17	6	9	.91	.96
Any Juvenile Arrest for Sex Offence & Adult Conviction for Separate Sex Offence		98	2			.66	.80
Rate of Sex Offending		84	16			.73	.85
Ever sentenced for a non-contact sexual offence		76	24			.89	.94
Any male victim		84	16			.85	.92
Two or more victims under age 12, one unrelated to offender.		76	24			.64	.78
Any unrelated victims		25	75			.70	.82
Any stranger victims		61	39			.68	.81
Any previous charges or convictions		31	35	28	6	.88	.93
Any community violation		54	46			.68	.81
Years free prior to index sexual offence		61	39			.55	.71
Ever previously sentenced for non-sexual violence		76	24			.87	.93

Note. $N = 100$.

Table G5 ***Frequency and Interrater Reliability of SORAG Ratings, Individual Risk Factors***

SORAG Items	Score Frequency											ICC_1	ICC_2
	-5	-3	-2	-1	0	1	2	3	4	5	6		
Biological parent			48					52				.78	.88
Elementary School					59		24			17		.67	.80
History alcohol problems				29	30	17	24					.85	.92
Marital status			82			18						.80	.89
Criminal history nonviolent				47	23			30				.81	.89
Criminal history violent				53	9						38	.66	.79
Number previous SO convictions				77		14				9		.69	.82
History of SO girls under 14					30				70			.59	.74
Failure on prior conditional release					53			47				.63	.78
Age index offence	38		24	18	4		16					.93	.96
DSM personality disorder			69					31				.69	.82
DSM schizophrenia		2				98						.49	.66
PCL score												1.00	1.00

Note. *N* = 100

.

Table G6 Correlation (r) Between RSVP Relevance Total and Domain Scores and SVR-20 Total and Domain Scores

RSVP Total & Domain,	SVR-20 Total Scores	History of sexual offences	Psychological adjustment	Social adjustment	Future plans
RSVP Relevance					
Total scores	.93***	.72***	.73***	.67***	.63***
History of sexual violence	.57***	.81***	.39***	.19	.23*
Psychological adjustment	.70***	.54***	.52***	.52***	.49***
Social adjustment	.70***	.36***	.52***	.71***	.49***
Manageability	.71***	.38***	.41***	.70***	.77***
Mental Disorder	.77***	.44***	.86***	.52***	.37***

Note. $N = 100$.

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

Table G7 Correlation (r) Between RSVP Presence Risk Factors and Static-99R, Static-2002R, and SORAG Total Scores and Risk Categories

RSVP Presence Items	Stat99R Total	Stat99R Risk Cat.	Stat2002R Total	Stat2002R Risk Cat	SORAG Total	SORAG Risk Cat.
Chronicity of sexual violence	-.00	.06	.04	.07	-.19	-.11
Diversity of sexual violence	.39***	.40***	.44***	.47***	.35***	.38***
Escalation of sexual violence	.02	.05	.03	.04	.08	.07
Physical coercion	.15	.15	.12	.09	.38***	.38***
Psychological coercion	-.31**	-.30**	-.27**	-.29**	-.19	-.18
Extreme minimization	.10	.16	.11	.11	.22*	.21*
Attitudes	.19	.13	.18	.12	.29**	.38**
Problems with self-awareness	.04	.09	.03	.03	.04	.05
Problems with stress	.18	.20	.22*	.22*	.25*	.26*
Problems from child abuse	.15	.09	.15	.08	.38***	.37***
Sexual deviance	.27**	.35***	.30**	.36***	.06	.09
Psychopathic personality	.44***	.41***	.42***	.39***	.65***	.66***
Major mental illness	.09	.11	.11	.14	.02	.04
Problems with substance use	.08	-.00	.10	.06	.39***	.37***
Violent or suicidal ideation	.19	.14	.19	.18	.36***	.25***
Intimate relationships	.19	.12	.14	.04	.25*	.24*
Non-intimate relationships	-.03	-.08	-.01	-.03	.20*	.20*
Problems with employment	.34**	.25*	.31**	.21*	.60***	.58***
Nonsexual criminality	.29**	.20*	.28**	.22*	.59***	.56***
Problems with planning	.33**	.32**	.38***	.32**	.47***	.44***
Problems with treatment	.38***	.37***	.42***	.39***	.53***	.51***
Problems with supervision	.52***	.51***	.60***	.61***	.68***	.65***

Note. Physical coercion = Physical coercion in sexual violence, Psychological coercion = Psychological coercion in sexual violence, Attitudes = Attitudes that support or condone sexual violence, Problems with stress = Problems with stress and coping, Problems from child abuse = Problems resulting from child abuse, Psychopathic personality = Psychopathic personality disorder, Intimate relationships = Problems with intimate relationships, Non-intimate relationships = Problems with non-intimate relationships. Stat99 Total = Static99R total scores, Stat99R Risk Cat = Static-99R risk category, Stat2002R Total = Static-2002R total scores, Stat2002R Risk Cat = Static-2002R risk category, SORAG Total = SORAG total scores, SORAG Risk Cat = SORAG risk category

$N = 100$, *$p < .05$, **$p < .01$, ***$p < .001$

Table G8 Correlation (r) Between RSVP Relevance Risk Factors and Static-99R, Static-2002R, and SORAG Total Scores and Risk Categories

RSVP Relevance Items	Stat99R Total	Stat99R Risk Cat.	Stat2002R Total	Stat2002R Risk Cat	SORAG Total	SORAG Risk Cat.
Chronicity of sexual violence	.11	.15	.18	.18	-.03	-.01
Diversity of sexual violence	.40***	.40***	.45***	.48***	.38***	.40***
Escalation of sexual violence	.15	.15	.17	.16	.20*	.19
Physical coercion	.16	.17	.12	.10	.40***	.39***
Psychological coercion	-.25*	-.24*	-.22*	-.23*	-.12	-.11
Extreme minimization	.00	.07	.01	.04	.20*	.21*
Attitudes	.20*	.15	.19	.14	.30**	.30**
Problems with self awareness	.02	.07	.00	-.01	.05	.04
Problems with stress	.19	.22*	.23*	.22*	.25*	.26*
Problems from child abuse	.26**	.19*	.22*	.15	.48***	.48***
Sexual deviance	.30**	.35***	.34**	.36***	.10	.13
Psychopathic personality	.44***	.41***	.42***	.38***	.65***	.66***
Major mental illness	.11	.11	.13	.16	.02	.04
Problems with substance use	.01	-.04	.02	-.01	.31**	.29***
Violent or suicidal ideation	.20*	.15	.17	.15	.28**	.27***
Intimate relationships	.15	.11	.10	.02	.19	.19
Non-intimate relationships	-.03	-.06	-.01	-.02	.16	.15
Problems with employment	.32**	.22*	.31**	.19	.52***	.50***
Nonsexual criminality	.22*	.16	.21*	.15	.48***	.46***
Problems with planning	.33**	.30**	.37***	.30**	.45***	.41***
Problems with treatment	.38***	.36***	.41***	.39***	.52***	.50***
Problems with supervision	.54***	.53***	.58**	.59***	.69***	.67***

Note. Physical coercion = Physical coercion in sexual violence, Psychological coercion = Psychological coercion in sexual violence, Attitudes = Attitudes that support or condone sexual violence, Problems with stress = Problems with stress and coping, Problems from child abuse = Problems resulting from child abuse, Psychopathic personality = Psychopathic personality disorder, Intimate relationships = Problems with intimate relationships, Non-intimate relationships = Problems with non-intimate relationships. Stat99 Total = Static99R total scores, Stat99R Risk Cat = Static-99R risk category, Stat2002R Total = Static-2002R total scores, Stat2002R Risk Cat = Static-2002R risk category, SORAG Total = SORAG total scores, SORAG Risk Cat = SORAG risk category.

$N = 100$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table G9 AUC for SVR-20 Ratings for Sexual Recidivism at 10-year follow up

				95% Confidence Interval	
SVR-20 Items	**AUC**	***SE***	***p***	**Lower**	**Upper**
Sexual Deviation	.62	.06	.063	.50	.74
Victim of Child Abuse	.53	.06	.632	.41	.65
Psychopathy	.63*	.06	.040	.51	.76
Major Mental Illness	.56	.06	.393	.43	.68
Substance Use Problem	.53	.06	.649	.40	.66
Suicidal/Homicidal Ideation	.56	.06	.378	.43	.68
Relationship Problems	.54	.06	.584	.41	.66
Employment Problems	.54	.06	.501	.42	.67
Past Non-sexual Violent Offences	.58	.07	.241	.45	.70
Past Non-violent Offences	.52	.06	.770	.40	.64
Past Supervision Failures	.68**	.06	.004	.57	.80
High Density Sex Offences	.51	.06	.882	.39	.63
Multiple Sex Offence Types	.62	.06	.054	.50	.75
Physical Harm to Victim(s)	.57	.07	.282	.44	.70
Uses Weapons/Threats of Death	.55	.07	.445	.42	.68
Escalation in Frequency/Severity	.51	.06	.885	.39	.63
Extreme Minimization/Denial	.49	.06	.906	.37	.62
Attitudes That Support/Condone	.59	.06	.176	.47	.71
Lack Realistic Plans	.61	.06	.077	.50	.73
Negative Attitude Toward Intervention	.53	.06	.613	.41	.66

The test result variable(s): Sexual Deviation, Victim of Child Abuse, Psychopathy, Major Mental Illness, Substance Use Problem, Suicidal/Homicidal Ideation, Relationship Problems, Employment Problems, Past Non-sexual Violent Offences, Past Non-violent Offences, Past Supervision Failures, High Density Sex Offences, Multiple Sex Offence Types, Physical Harm to Victim(s), Uses Weapons/Threats of Death, Escalation in Frequency/Severity , Extreme Minimization/Denial, Attitudes That Support/Condone, Lack Realistic Plans, Negative Attitude Towards Intervention has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Note. $N = 100$. *$p < .05$, **$p < .01$.

Table G10 AUC for Static-99R Risk Factors for Sexual Recidivism at 10-year follow up

				95% Confidence Interval	
Static-99R Items	**AUC**	***SE***	***p***	**Lower**	**Upper**
Prior sex offences	.63*	.07	.042	.50	.76
Prior sentencing dates	.56	.07	.325	.44	.69
Non-contact sex convictions	.53	.07	.693	.40	.65
Index non-sex violent conviction	.58	.07	.198	.45	.71
Prior non-sex violent conviction	.52	.06	.814	.39	.64
Any unrelated victims	.60	.06	.133	.48	.71
Any stranger victims	.59	.06	.161	.47	.71
Any male victims	.53	.07	.605	.41	.66
Young	.53	.07	.598	.41	.66
Ever lived with a partner	.50	.06	.949	.38	.63

The test result variable(s): Prior sex offences, Prior sentencing dates, Non-contact sex convictions, Index non-sexual violent conviction, Prior non-sexual violent conviction, Any unrelated victims, Any stranger victims, Any male victims, Young, Ever lived with... has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Note. $N = 100$. *$p < .05$, **$p < .01$.

Table G11 AUC for Static-2002R Risk Factors for Sexual Recidivism at 10-year follow up

				95% Confidence Interval	
Static-2002R Items	**AUC**	***SE***	***p***	**Lower**	**Upper**
Age at release	.55	.06	.427	.43	.67
Static 2002 item 2 total recoded	.63*	.07	.037	.51	.76
Non-contact sex-offending	.55	.07	.438	.42	.68
Any male victims	.53	.07	.605	.41	.66
2+ victims <12yrs, 1 unrelated	.62	.07	.056	.50	.75
Any unrelated victims	.60	.06	.106	.49	.72
Any stranger victims	.59	.06	.161	.47	.71
Static 2002 item 5 total recoded	.61	.06	.090	.49	.73

The test result variable(s): Age at release, Static 2002 item 2 total recoded, Non-contact sex-offending, Any male victims, 2+ victims <12yrs, 1 unrelated, Any unrelated victims, Any stranger victims, Static 2002 item 5 total recoded has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Table G12 AUC for SORAG Risk Factors for Sexual Recidivism at 10-year follow up

				95% Confidence Interval	
SORAG Items	**AUC**	***SE***	***p***	**Lower**	**Upper**
Biological parent	.52	.06	.727	.40	.65
Elementary School	.64*	.06	.025	.52	.76
History alcohol problems	.53	.06	.603	.41	.66
Marital status	.52	.07	.767	.39	.65
Criminal history nonviolent	.54	.06	.487	.42	.67
Criminal history violent	.59	.07	.148	.47	.72
Number previous SO convictions	.62	.07	.073	.49	.74
History of SO girls under 14	.54	.06	.518	.42	.66
Failure on prior conditional release	.70**	.06	.001	.59	.82
Age index offence	.59	.06	.143	.47	.71
DSM personality disorder	.60	.06	.128	.47	.72
DSM schizophrenia	.51	.06	.826	.39	.64
PCL score	.60	.06	.133	.48	.72

The test result variable(s): Biological parent, Elementary School, History alcohol problem, Marital status, Criminal history nonviolent, Criminal history violent, Number previous SO convictions, History of SO girls under 14, Failure on prior conditional release, Age index offence, DSM personality disorder, DSM schizophrenia, PCL score has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Note. $N = 100$. *$p < .05$, **$p < .01$.