



Kent Academic Repository

Raymond, Bea, Davis, Michael R., Ogloff, James R. P. and McEwan, Troy E.
(2024) *Convergent and predictive validity of the Sexual Violence Risk-20 and risk for sexual violence protocol with older sexual offenders. Psychology, Crime & Law*, 32 (5). pp. 834-855. ISSN 1068-316X.

Downloaded from

<https://kar.kent.ac.uk/115939/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1080/1068316X.2024.2401489>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

Additional information

Versions of research works

Versions of Record

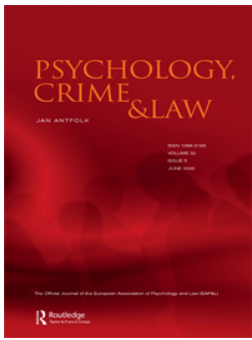
If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).



Convergent and predictive validity of the Sexual Violence Risk-20 and risk for sexual violence protocol with older sexual offenders

Bea Raymond, Michael R. Davis, James R. P. Ogloff & Troy E. McEwan

To cite this article: Bea Raymond, Michael R. Davis, James R. P. Ogloff & Troy E. McEwan (2026) Convergent and predictive validity of the Sexual Violence Risk-20 and risk for sexual violence protocol with older sexual offenders, *Psychology, Crime & Law*, 32:5, 834-855, DOI: [10.1080/1068316X.2024.2401489](https://doi.org/10.1080/1068316X.2024.2401489)

To link to this article: <https://doi.org/10.1080/1068316X.2024.2401489>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 16 Sep 2024.



[Submit your article to this journal](#)



Article views: 5444



[View related articles](#)



[View Crossmark data](#)



Citing articles: 3 [View citing articles](#)



Convergent and predictive validity of the Sexual Violence Risk-20 and risk for sexual violence protocol with older sexual offenders

Bea Raymond^{a,b}, Michael R. Davis^{a,c,d}, James R. P. Ogloff^a and Troy E. McEwan^{a,e}

^aCentre for Forensic Behavioural Science, Swinburne University of Technology and Forensicare, Melbourne, Australia; ^bVicpsychplus, Melbourne, Australia; ^cDepartment of Psychiatry, Monash University, Melbourne, Australia; ^dDepartment of Psychiatry, University of Melbourne, Melbourne, Australia; ^eCentre for Research and Education in Forensic Psychology University of Kent, Canterbury, England

ABSTRACT

Structured professional judgement (SPJ) guidelines for assessing risk of sexual offending have existed for more than two decades, yet the literature investigating the validity of these tools remains relatively limited. Recent research has shown that actuarial sexual offending risk assessment tools should be adjusted to reflect reduced risk as offenders age (Helmus et al., 2012), leading to questions about the validity of extant SPJ guidelines for older people who sexually offend. The current study investigated the reliability and predictive validity of the Sexual Violence Risk-20 (SVR-20) and the Risk for Sexual Violence Protocol (RSVP) with a sample of 95 Australian individuals aged 50 years or over who had sexually offended. Sixteen (16.84%) participants sexually recidivated over a mean follow-up period of 9.07 years. Contrary to results with younger offenders, structured judgements using the SVR-20 or RSVP did not significantly differentiate recidivists from non-recidivists in this sample (AUC = .57). However, the total scores for both instruments were predictive at particular time periods. Certain risk factors were potentially important and accounted for unique variance that was not accounted for by actuarial instruments. The implications and limitations of this study are discussed, as well as recommendations for future research and practice.

ARTICLE HISTORY

Received 31 October 2023
Accepted 28 August 2024

KEYWORDS

aged; SVR-20; RSVP;
recidivism; sexual

Sexual offending and its associated harms are a source of substantial community concern. For more than two decades, this concern has led to increased focus on the need to effectively assess and manage the risk presented by those who sexually offend (Harris et al., 2010). However, the inherent heterogeneity of these offenders, the range of their sexual behaviours, and the variety of potential victims, means that assessing the risk of future sexual offending is not straightforward. In the 1990s, researchers began to identify a range of risk factors for repeat sexual offending, which has informed the development of

CONTACT Bea Raymond  bea.raymond@hotmail.com  Vicpsychplus, PO Box 5007, Keon Park Victoria 3073, Melbourne, Australia

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

different tools to assist in assessing this risk. The tools with the most research support are actuarial or statistically-derived instruments, which have shown moderate to good predictive validity for sexual reoffending (Hanson & Morton-Bourgon, 2009). However, these types of tools tend to be reliant on static (i.e. relatively unchanging) risk factors, making them less useful for informing the development of formulations and risk management plans (Beech et al., 2016; Rettenberger et al., 2009a). In contrast, structured professional judgement (SPJ) risk assessment guidelines are rationally derived from the wider literature and consider both static and dynamic (i.e. potentially changeable) factors. It has been suggested that this makes them more useful for guiding treatment and risk management planning (Craig et al., 2004; Holoyda & Newman, 2016; Ogloff & Davis, 2020). Given the potential additional benefit of SPJ guidelines risk assessment to help measure change and inform risk management strategies, it is essential to evaluate the reliability and validity of such measures.

More recently, there has been a growing awareness that the most commonly used sex offender risk assessment tools may not show comparable predictive validity in particular populations. For example, research in groups of offenders such as adolescents, females, and those with a cognitive impairment, has led to the development and validation of separate instruments for these groups (Blackley & Bartels, 2018; Boer et al., 2004; Cortoni et al., 2010; Cortoni et al., 2017; Fedoroff et al., 2016; Kelley & Thornton, 2015; Lindsay & Taylor, 2010; Ralston & Epperson, 2013; Stevens et al., 2013; Worling, 2004; Worling et al., 2012; Worling & Curwen, 2000). Despite a range of characteristics differentiating older men (i.e. 50 years and older) from other sexual perpetrators, limited research exists that investigates the efficacy of existing instruments with such offenders (Barbaree et al., 2008; Hanson, 2002, 2006; Nicholaichuk et al., 2014; Raymond et al., 2020).

The need for further investigation is necessary as the population of sexual and non-sexual offenders, who are aged, continues to grow around the industrialised world. In Australia, the fastest-growing group within prisons are those aged over 65 years (Baidawi et al., 2011; Shaw et al., 2018). This is partly a consequence of increased convictions and concurrent sentencing of older males for historical sexual offences (Chua et al., 2018; Shaw et al., 2018). The increasing number of older males who sexually offend that are in prisons or involved with probation services, makes it particularly important that commonly used risk assessment tools are shown to be valid with this group (also see Australian Institute of Health and Welfare, (2020).

Older men and risk of sexual offending

Generally, conviction and recidivism rates decrease as an offender ages, regardless of offence type or history (Bengtson et al., 2019; Nicholaichuk et al., 2014; Rakes et al., 2018). Research investigating age and sexual offending has shown that advanced age is generally a protective factor for both perpetrating initial sexual offences and sexual recidivism (Nicholaichuk et al., 2014; Olver et al., 2012; Rettenberger et al., 2013). Factors that are considered to effect risk of recidivism for the older sex offender population have largely been gleaned from the research on aging non-offenders. Indeed, researchers have attributed the protective role of age to the normal effects of aging, such as decreased libido, lowered production of testosterone, and often greater limitation to accessing potential victims as a result of mobility issues. The assumption is that the

biological changes that occur naturally due to age, contribute to a reduction in the risk of recidivism (Hanson, 2001; Nicholaichuk et al., 2014; Rodriguez et al., 2017).

The protective effect of being older has been incorporated into some existing sexual offender risk assessment tools, such as the Static-99R (Phenix et al., 2016), the most recent edition of the Static-99 family of instruments. The Static-99R was developed following research that showed the Static-99 overestimated risk of older offenders. The age weights included for the Static-99R provide different scores based on the offender's age at the time of release (Helmus et al., 2012). Younger offenders receive a higher score on the age item than older offenders. A recent Australian study by Reeves and colleagues (2017) investigated the effects of this age-correction in a sample of men aged 18–78 with a history of sexual offending. They found both the original Static-99 (Hanson & Thornton, 2000; Harris et al., 2003) and the revised version were similar in their ability to discriminate between recidivists and non-recidivists. However, further analyses by Raymond and colleagues (2020) in the sub-sample of offenders over the age of 50 years, found weaker predictive validity in those over the age of 60 years, irrespective of the age correction. Other actuarial tools such as the Violence Risk Appraisal Guide family of instruments (VRAG/SORAG/VRAG-R; Harris et al., 2015; Quinsey et al., 1998) also incorporate age in a somewhat different manner using factors based on age at the time of the index offending. Harris and Rice (2007) argue that age at release had little impact on recidivism prediction while age at first offence improved predictive validity. In contrast, Barbaree and colleagues (2008) found that age at first offence was static and unchanging and does not account for the protective aspects of the aging process. This research demonstrates that the authors of actuarial instruments have moved to take into account the apparent protective effects of age, however the same is not true of SPJ risk assessment guidelines intended to assess risk of sexual recidivism.

The validity of the SVR-20 and RSVP

Unlike the strictly mechanical approach of the actuarial tools, SPJ risk assessment guidelines provide a degree of discretion to the risk assessment task. The convergent approach of utilising an actuarial tool in conjunction with SPJ risk assessment guidelines has been viewed as best practice by some authors in relation to assessing risk for sexual recidivism (e.g. Boer, 2006; Boer & Hart, 2009; Rettenberger et al., 2009b). Indeed, the risk factors contained in SPJ risk assessment guidelines are not intended to be exhaustive, and there is an expectation that the assessor will exercise their knowledge and discretion when applying these tools (Hart & Boer, 2009). This may be one way that issues associated with increased offender age can be considered when making final risk judgements.

Two of the most commonly used SPJ risk assessment guidelines for sexual violence are the Sexual Violence Risk-20 (SVR-20; Boer et al., 1997) and the Risk for Sexual Violence Protocol (RSVP; Hart et al., 2003). Both tools were rationally developed based on risk factors known or believed to be associated with sexual recidivism at the time of development (Cheng et al., 2019; Kelley et al., 2018). Of the two, the SVR-20 has been subject to more validation research, with studies in Switzerland, Germany, Austria, United Kingdom, Canada, and the Netherlands, suggesting the tool has a moderate to strong ability to differentiate between sexual recidivists and non-recidivists (AUCs generally ranging from .63 to .83 across studies; Barbaree et al., 2008; de Vogel et al., 2004;

Kanters et al., 2017; Rettenberger & Hucker, 2011; Rettenberger et al., 2011; Rettenberger et al., 2009b). Indeed, a meta-analysis by Hanson and Morton-Bourgon (2009) found that structured judgements made on the basis of the SVR-20 guidelines were the most accurate of all risk assessment approaches for sexual offenders, using three of the studies that had a combined sample of 245 ($d = 1.23$; $AUC = .78$). There have been limited studies that have found smaller relationships between total 'scores' on the instrument and sexual recidivism that have sometimes not been statistically significant (e.g. see Kelley & Thornton, 2015).

Few published studies have specifically evaluated the reliability and validity of the RSVP, although given the very strong correlation with total scores on the SVR-20 ($r = .97$), they have been considered parallel or equivalent forms (Hart & Boer, 2009). Inter-rater reliability of the RSVP has been discussed by Sutherland and colleagues (2012) and Vargen and colleagues (2020). Both studies concluded that the RSVP is reliable between raters, with similar overall case prioritisation judgements (ICC ranged between .62 to .91 for overall case prioritisation), although item-level agreement was more variable. Sutherland and colleagues' findings suggested that reliability improved with assessor experience. In a more recent study by Sea and Hart (2021) interrater reliability was found to be excellent when users of the RSVP were specifically trained in the use of the guidelines.

Vargen and colleagues (2020) conducted a study of 100 Canadian offenders who had participated in sex offender treatment between 2002 and 2004. The RSVP was scored from file review and offenders were followed-up for an average of 9.5 years using law enforcement data (i.e. new police contacts, charges and convictions). Recidivism rates for police contacts increased across case prioritisation categories (17% in the low category, 32% in moderate, and 62% in the high category). Furthermore, those judged to be of 'high' case priority had 7.6 times greater odds of reoffending than those judged low risk (there was no significant difference between odds of reoffending in the moderate and low risk groups). No Receiver Operating Characteristic (ROC) analysis was reported but was easily discernible from the published data to enable comparison with other risk assessment studies ($AUC = .67$). Vargen and colleagues reported moderate to strong correlations between RSVP total scores and case prioritisation judgements. Convergent validity was also found with total scores on the SORAG and Static-99R. The age of the participants ranged between 19 and 77 years ($M = 41$, $SD = 12.41$); however, there was no further information indicating the portion of the sample that could be categorised as aged. While age did not significantly predict reoffending when examined alone, it tended to significantly improve fit when used in models with other risk instruments and was controlled in the final analyses.

A recent study by Davis et al. (in submission) investigated the predictive validity of the RSVP in a sample of 115 adults aged between 18 and 46 years of which 13.04% were assessed as falling in the low category, 47.82% in moderate, and 39.13% in the high category. Within the sample, 24 engaged in sexual reoffending, all of whom had been assessed as moderate- or high-risk. When evaluated on the ability to distinguish between recidivists and non-recidivists, total RSVP scores proved more effective than the assessor's case prioritisation judgement in this retrospective study based on file material (RSVP Lifetime Total Score $AUC = .73$; Case Prioritisation $AUC = .68$).

Using a somewhat different design to the two prior studies, Darjee and colleagues (2016) investigated the predictive validity of the RSVP in a sample of 109 Scottish

males, aged between 17 and 68 years, referred to a forensic mental health service for sexual offending or harmful sexual behaviour (Darjee et al., 2016). Other tools utilised included the Risk Matrix 2000 (RM2000; Thornton et al., 2003) and the Hare Psychopathy Checklist-Revised (PCL-R; Hare, 2003). This was a genuinely prospective study whereby recidivism data were obtained for a brief follow-up period ($M = 3.25$ years, $SD = 1.77$). Almost half of the sample (47; 43.1%) committed another offence (allegation, charge or conviction) and approximately 38 (33.9%) re-offended sexually. Each participant was assessed by two assessors on a minimum of two occasions. This study was different to other research on SPJ guidelines risk assessment as it was conducted in the field while other research has relied upon file information to score the RSVP, meaning the researchers were able to assess how well the RSVP worked in real-world settings. The results indicated that the RSVP had convergent validity with an actuarial instrument (RM2000; Thornton et al., 2003) as well as a measure of psychopathy (PCL-R; Hare, 2003), and behavioural scales of sexual deviance. However, the results also showed that the ability of the RSVP to differentiate between sexual recidivists and non-recidivists was not statistically significant ($AUC = .59$, 95% CI [.48, .71]). The authors suggested that this result may have been due to the effects of risk management, given it was not possible to undertake prospective research without risk management guided by the risk assessments undertaken. The sample was particularly high risk and not representative of the wider population of those who sexually offend, which may have affected the performance of the judgements. The performance of the RSVP case prioritisation judgement with the subsample of older participants was not specifically considered in this study.

Older males and SPJ guidelines risk assessment of sexual offending

Incorporating age at release, victim type such as child or adult (accessibility), and the effects of physical decline (including physical capacity, mental, and neurological changes) when scoring particular risk factors of the SVR-20 and RSVP, may assist evaluating recidivism risk (Barbaree et al., 2008). However, age does not explicitly form part of the tools' item descriptions or risk formulation guidance, making this approach reliant on the clinician's knowledge of how age may affect sexual behaviour. For example, a clinician may not have thorough knowledge of the age effects on cognitive and physical functioning, and age-related psychological factors such as depression associated with reduced function, lack of independence and loss of relationships. An additional potential issue affecting risk assessment in this group is an assumption that as people age their sexual interest and need for intimacy universally decreases as professionals and those in the community may view old age and intimacy as mutually exclusive (Bauer et al., 2009).

Ironically, factors that may play a different role in sexual offending risk among older adults are similar to those observed to differ among juveniles, such as social isolation and biological changes (Gunby & Woodhams, 2010; Ozkan et al., 2020). It is possible that like juveniles who sexually offend, the older offender should be assessed using modified or other specific risk assessment tools. At the very least, an assessor should have the knowledge and skill to recognise and ascertain the effect of age-specific considerations to undertake a valid risk assessment, which includes understanding that age is not always protective. As Nicholaichuk and colleagues (2014) highlighted, there is a small group of older men who sexually offend for the first time or continue to

offend over time. This small subsample may represent the portion of men who have enduring sexually deviant interests or have a considerably elevated number of risk factors present such as historical psychosocial problems, sexual and antisocial factors.

As an alternative to relying on assessors' knowledge of age effects, Barbaree and colleagues (2008) investigated whether the predictive validity of the SVR-20 could be improved using an age correction on the tool's total score (similar to the adjustments applied to actuarial tools). A sample of 468 Canadian sexual offenders aged between 21 and 83 years-old were followed-up for an average of five years. The authors concluded that addition of age-at-release had little effect on the predictive validity of the overall total score of the SVR-20. A study conducted by Olver et al. (2020) reviewed the use of age at release with the Violence Risk Scale-Sexual Offences instrument (VRS-SO; Wong et al., 2017) and the Static-99. The sample consisted of 1,287 participants who were sourced from four forensic services, including those over 50 ($n = 158$) and those 60 and over ($n = 77$). The study investigated the effect of age-related decline of criminality, the discrimination across different age groups, the changes and calibration of the VRS-SO and Static-99R. These two tools were combined by the Static-99R items replacing the VRS-SO static items and it was shown that the age at release item did not add to the integrated model (unsurprisingly, given both tools account for age themselves). It was noted that those over the age of 60 years tended to be overpredicted in regard to risk. Nonetheless, this study also found the VRS-SO's post-treatment dynamic facet score and the instrument's total post-treatment score to be predictive for men in their 50s, and all factors/facets of the VRS-SO were found to be predictive of sexual recidivism for men older than 60.

Given the increased attention to sexual offending and risk assessment among older offenders, and the lack of evaluation studies of SPJ risk assessment guidelines in this group to date, the current study undertook to investigate the validity of the SVR-20 and the RSVP in a sample of older men with a history of sexual offending. The aim was to examine inter-rater reliability, internal consistency, convergent validity with other risk assessment tools including the Static-99R, and predictive validity.

Method

Participants

The SVR-20 and RSVP were scored from the clinical files of 95 men aged 50 years or older (mean age = 59.98, $SD = 7.08$) at the time of assessment, all with a known history of sexual offending. Due to sample size, there was no differentiation between those who were convicted for historical offences, persistent offenders, or offended after the age of 50. Participants had all been assessed by a forensic psychologist or psychiatrist at a community forensic mental health service in Melbourne, Australia between 1990 and 2011 after being referred by courts ($n = 20$, 21.1%), probation services ($n = 64$, 67.4%), or self-referral ($n = 3$, 3.2%) (seven files were missing referral information). Participants were identified from a larger dataset of all 621 men assessed for sexual offences at the service during that time (see Reeves et al., 2017). Of that sample, 118 were aged 50 years or older, but 23 were excluded due to insufficient file information to allow scoring of the SVR-20 and RSVP. Excluded cases tended to be from the earlier years of data collection, potentially because the level of detail in file information was routinely less in older files.

Information about the predictive performance of the Static-99 and Static-99R in the total sample of older men ($n = 118$) has previously been published by Raymond and colleagues (2020), and are not further investigated in this study. Rather, these two tools are used largely as measures of convergent validity in the current study.

Measures

SVR-20. The SVR-20 (Boer et al., 1997) was developed in 1997 and used for the current research. A revised edition became available in 2017, which resulted in updating the list of risk factors, drawing on research at the time that identified dynamic and static risk factors associated with sexual reoffending. Sexual violence was defined as ‘actual, attempted, or threatened sexual contact with a person who is non-consenting or unable to give consent’ (Boer et al., 1997, p. 9). The clinician is guided through the process of making an overall summary risk rating that culminates into low, moderate, or high risk of further sexual violence. The 1997 version, used in this research, consists of 20 risk factors grouped into three domains: Psychosocial Adjustment consisting of 11 items; History of Sexual Offences consisting of seven items; and Future Plans covering the final two items. For research purposes, the Psychosocial Adjustment domain has been further divided into Psychological Adjustment (six items) and Social Adjustment (five items). For each item, it must be determined whether the factor has been present at any time within the client’s life. Each factor is scored on a three-point ordinal scale ($N = no\ evidence$, $? = possible/partial\ evidence$, and $Y = definite\ presence$). This is translated into a numerical scale for research purposes ($N = 0$, $? = 1$, $Y = 2$). After scoring the presence of all 20 risk factors, items are scored in relation to any change during the preceding year ($+ = exacerbation$, $0 = not\ present$, $- = amelioration$). The user then makes a judgement about the risk of future sexual violence being low, moderate, or high; with relatively little guidance about how to draw such a conclusion. The psychopathy risk factor was not scored during the current study due to lack of available information using the pseudo-prospective file review design. Accordingly, only 19 of the 20 factors were used to inform the risk judgement. It should be noted that the summary risk ratings were made based on the file information available, with particular attention paid to the number, nature, and perceived relevance of the risk factors to each man’s harmful sexual behaviour.

RSVP. The RSVP (Hart et al., 2003) was developed using a similar approach to the SVR-20 and can be considered a slightly expanded parallel form. The RSVP comprises 22 risk factors divided into five domains: History of Sexual Violence (five items); Psychological Adjustment (five items); Mental Disorder (five items); Social Adjustment (four items); and Manageability (three items). Each item is scored using the same three-point ordinal scale as the SVR-20. However, each risk factor is initially coded twice. Past presence refers to the individual’s life up until 12 months before the assessment. Recent presence refers exclusively to the previous 12 months. For this study, each risk factor rating was converted to a numerical scale ($0 = not\ present$, $1 = partial\ presence$, and $2 = present$). Life-time scores were then calculated by ascertaining the highest score between the past and present ratings. A third rating of future relevance is also applied to each item in the RSVP and should be used to help inform risk scenarios and overall risk judgements.

The RSVP manual provides more guidance for assessors than the SVR-20. The step-by-step guidelines assist the clinician to develop risk scenarios and make strategies for

managing risk and summary judgements. The latter includes case prioritisation, the degree of harm should the risk materialise, and whether risk of reoffending is imminent. Although judgements about imminence and severity of future violence are also typically made, only case prioritisation was considered in analyses in this study. As was the case with the SVR-20, the psychopathy item was not scored due to lack of necessary information. As such, only 21 of the 22 factors were used when developing case prioritisation judgements. While severity and imminence were not formulated, case prioritisation judgements were made based on the totality of information available, and coded as 0 = *Low*, 1 = *Moderate*, 2 = *High*.

Static-99 and Static-99R. The Static-99 (Harris et al., 2003) and Static-99R (Phenix et al., 2016) are actuarial risk assessment tools that both contain 10 historical/static risk factors. The items on the Static-99 are scored dichotomously except for one (Prior Sentencing Dates) which is coded on an ordinal scale ranging from zero to three. Possible scores range from zero to 12. The total score is used to mechanically place an individual into one of four nominal risk categories (0–1 = low; 2–3 = moderate-low risk; 4– = moderate-high risk; 6+ = high). The Static-99R is scored identically to the Static-99 except for the age item. Rather than the dichotomous split at 25 years that is used for the Static-99, the Static-99R includes four possible age scores based on time exposed to risk (–3 = 60+; –1 = 40–59.9; 0 = 35–39.5; 1 = 18–34.9). The highest score that can be obtained is still 12, but the lowest score that can be obtained is now –3. The total score is used to mechanically place an individual into one of five nominal risk categories (–3 to –2 = very low risk; –1–0 = below average risk; 1–3 = average risk; 4–5 = above average risk; and 6+ = well above average risk).

Procedure

A pseudo-prospective design was used with the first author accessing file data and coding from information available at the time of the original assessment, and each case then followed up using administrative data. The primary author is a highly experienced forensic psychologist (20+ years of practice), formally trained in the use of actuarial and SPJ risk assessments. She rated all tools blind to any previous risk assessments undertaken by the original assessing clinician, and to follow-up data about reoffending. File information was initially used to score the Static-99 and Static-99R in a randomised order (Raymond et al., 2020). At the completion of that study, and more than one year later, the same file data was accessed by the same rater and information was used to score the RSVP and SVR-20. The RSVP and SVR-20 were scored blind to the Static 99 and Static-99R data and completed in a randomised order. The RSVP and SVR-20 were never scored straight after one another from the same file, with scoring of another case always interspersed. Time between scoring the two sets of guidelines in the same case ranged from on the same day to approximately one week apart. This took place over approximately four-months until all identified files were alternately scored using the SVR-20 and RSVP. Where file information permitted, details of sexual, other offending and victim data were recorded.

Risk assessments completed by the original assessing clinician were used to ascertain inter-rater reliability, although the assessor in this study was blind to these at the time of completing file reviews. The RSVP had been used 12 times between February 2004 and

June 2010. The SVR-20 had been used 16 times between March 1998 and February 2003. Recidivism data were collated from the dataset developed by Reeves and colleagues (2017), which included subsequent convictions for sexual offences from the date of the original assessment until 2011. Follow-up time controlled for death and time incarcerated and ranged from 0.25 years to 18.87 years, with a mean follow up time of 9.07 years ($Mdn = 9.92$, $IQR = 4.20$ – 13.66 years). Human research ethics approval was granted by the relevant Swinburne University committees.

The decision to remove the psychopathy factor for both the SVR-20 and RSVP was based on insufficient file information to score the PCL-R. Review of the files where the original assessor scored either the SVR-20 or RSVP, showed no indication of psychopathy. The decision was therefore made to remove this factor from the data analysis.

Statistical analysis

Statistical analyses used the Statistical Package for the Social Sciences (SPSS), version 23. Medcalc was utilised to determine the statistical significance of differences in ROC curves. Inter-rater Reliability (IRR) was assessed using intraclass correlation coefficients (ICC_2 ; two-way, random effects model, single ratings, absolute agreement method) (Hallgren, 2012; McHugh, 2012). There are several published recommendations for interpreting ICC values. Fleiss (1981) suggested that for a single measure ICC, .75 and above can be considered 'excellent', .60 to .74 are 'good', .40 to .59 are 'moderate', and below .40 are 'poor'. This differs only slightly from Landis and Koch (1977) who proposed that above .81 was 'almost perfect', .61 to .80 was 'substantial', .41 to .60 was 'moderate', .21 to .40 was 'fair', .00 to .20 was 'slight' and below zero was 'poor'.

The Pearson product-moment correlation coefficient (r) was used to ascertain the convergent validity between the RSVP and SVR-20. Kaplan-Meier survival curves and log-rank analysis were used to determine significant differences in time to reoffending and the proportion of recidivism in each risk category. The ability of each instruments' case prioritisation/risk judgement to discriminate between recidivists and non-recidivists was evaluated using Receiver Operating Characteristic (ROC) analysis, and associated areas under the curve (AUC) (Singh et al., 2015). An AUC value of .50 indicates discrimination between groups no better than chance, while values greater than .70 indicate a strong discriminatory ability (Youngstrom, 2013). The sensitivity and specificity were calculated for Low versus Moderate/High categories to identify the proportion of recidivists who would be accurately identified as being at greater than low risk. Finally, it should be noted that other than the assessed risk category the factors were also summed to provide a total. These totals were used in the statistical analysis without weighting or identifying relevance to assessed risk level.

Results

The following describes the data obtained from the 95 participants selected for this study (noting that it is a sub-sample of the sample reported in Raymond et al. (2020)). The 95 participants ranged in age from 50 to 78 years old ($M = 59.98$, $SD = 7.08$). Seventy (73.69%) were married, separated, divorced or widowed at time of assessment, indicating a history of relationships, while 21 (22.11%) had never married. The offending that led to

assessment (*index offences*) were perpetrated against 23 males (24.21%) and 61 females (64.21%); seven had both male and female victims (7.37%), and four of the victims were of unknown sex (4.21%). The index offences mostly involved child victims; two victims were aged over 18 years (2.11%), nine between 12 and 17 years (9.47%), and 83 were up to 12 years old (87.37%) at the time of the offending. There were 35 (36.84%) victims who were related to the perpetrator, 39 (41.05%) were known to the offender and 15 (15.8%) were unknown to the offender. There were insufficient details for 6 (6.32%) of the victims to determine prior relationship. Forty-seven participants (49.47%) had sexual offences recorded prior to the index offence. No index offences were perpetrated in concert with others.

Participants' index offending was divided into contact offences ($n = 85$, 89.47%) and non-contact offences ($n = 10$, 10.53%). The contact offences included incest ($n = 11$, 11.58%), indecent act ($n = 25$, 26.32%), sexual penetration ($n = 14$, 14.74%), rape ($n = 3$, 3.16%), indecent assault ($n = 28$, 29.47%), and gross indecency ($n = 4$, 4.21%). The non-contact offences included wilful and obscene exposure ($n = 7$, 7.37%) and make/produce/possess child exploitation material and offensive behaviour in fewer than three cases each.

Sixteen participants (16.84%) were convicted of further sexual offences during the follow-up period, with time to reoffending varying between one month and 15.35 years ($M = 3.13$, $SD = 2.55$) after their first assessment leading to study inclusion (controlling for time incarcerated and death). It is important to note that, unsurprisingly given the nature of the sample, age was not related to sexual recidivism, $r_{pb}(93) = -.085$, $p = .41$. Eleven participants (12.63%) died during the follow-up period.

A total of 40 offences were committed after the index offences including incest ($n = 3$, 7.5%); wilful and obscene exposure ($n = 4$, 10%); sexual penetration ($n = 3$, 7.5%); rape ($n = 3$, 7.5%); indecent act ($n = 11$, 27.5%); loitering ($n = 11$, 27.5%); while fewer than three participants reoffended with any of indecent assault, child exploitation material, stalking or offensive behaviour. The victims were predominantly children aged between 6 and 11 years old ($n = 8$, 50%), children aged between 12 and 17 years old ($n = 7$, 43.75%). The victims included three (18.75%) males and eight (50%) females, and five (31.25%) offended against both males and females. Of the recidivists, 6 (37.5%) victims were known for longer than 24 h, 5 (31.25%) were related, 3 (18.75%) were unknown to the offender and there was insufficient information regarding 2 (12.5%) victims. Fourteen of the 16 recidivists had committed prior sexual offences ($n = 7$, 87.5%).

Risk assessment results

Table 1 presents the proportion of participants in each risk category on the RSVP and SVR-20, and rates of sexual recidivism in each category, with information about the Static tools also provided for comparison.

Inter-rater reliability

Inter-rater reliability for risk judgements and case prioritisations between the file review and original clinicians' ratings was in the 'excellent' (Fleiss, 1981) or 'almost perfect' (Landis & Koch, 1977) range for both the RSVP case prioritisation ($ICC_2 = 1.00$) and the

Table 1. SVR-20, RSVP, Static-99 and Static-99R risk judgement distribution and rates of sexual recidivism over an average 8.83 year follow-up ($N = 95$).

Categories	N (% of sample)	Sexual recidivism N (% within risk category)
RSVP		
Low	21 (22.1)	3 (14.3)
Moderate	43 (45.3)	6 (14.0)
High	31 (32.6)	7 (22.6)
SVR-20		
Low	23 (24.2)	3 (13.0)
Moderate	45 (47.4)	7 (15.6)
High	27 (28.4)	6 (22.2)
Static-99		
Low	27 (28.7)	2 (7.4)
Moderate – Low	35 (37.2)	3 (8.6)
Moderate – High	19 (20.2)	4 (21.1)
High	13 (13.8)	6 (46.2)
Static-99R		
Very Low	13 (13.8)	2 (15.4)
Below Average	31 (33.0)	0 (0.0)
Average	37 (39.4)	5 (13.5)
Above Average	11 (11.7)	6 (54.5)
Well Above Average	2 (2.1)	2 (100)

SVR-20 risk judgement ($ICC_2 = .97$, 95% CI [.91, .99]). Excellent inter-rater reliability for the Static and Static-99R scores were previously reported by Raymond and colleagues (2020) based on 30 cases from the larger parent sample (Static-99 $ICC_2 = .91$, 95% CI [.81, .96]; Static-99R $ICC_2 = .93$, 95% CI [.84, .96]). Each risk factor was also assessed for inter-rater reliability and the results indicated excellent to perfect reliability between the assessors (RSVP individual factor ICC_2 ranged between .76 and 1; SVR-20 individual factor ICC_2 ranged between .9 and 1).

Convergent validity

Table 2 shows the convergent validity of the various risk categories arising from the tools. The current rater's structured judgements made using the SVR-20 and RSVP differed in eight cases. In two cases there was a low-moderate discrepancy and in six cases a moderate-high discrepancy. Nonetheless, the correlation between the structured risk rating of the SVR-20 and the case prioritisation level of the RSVP was strong and statistically significant, as were the categories on the two variants of the Static-99. The RSVP categorisation was significantly correlated with the Static-99R but not the Static-99, though the size of the correlations was similar and substantially smaller than the correlations between SPJ risk assessment guidelines. Similar results were found between the SVR-20 risk

Table 2. Pearson's correlations (r) between risk categories on the different tools.

	SVR-20 Categories	RSVP Categories	Static-99 Categories	Static-99R Categories
SVR-20	-			
RSVP	.92**	-		
Static-99	.25*	.19	-	
Static-99R	.26**	.21*	.84**	-

Note: $N = 95$. * = $p < .05$; ** $p < 0.01$.

judgements and the risk categories for the versions of the Static-99, though correlations were significant for all relationships.

SVR-20 total scores ranged from 0 to 30 ($M = 10.76$, $SD = 6.04$) and on the RSVP, participants' lifetime scores ranged from 2 to 74 ($M = 30.83$, $SD = 14.67$). On the Static-99, the scores ranged from 0 to 9 ($Mdn = 3$, $IQR = 1-4$) while scores on the Static-99R ranged between -3 and 8 ($Mdn = 1$, $IQR = -1-2.25$). Table 3 shows the correlations between total scores of each of the tools. Total scores derived from the SPJ risk assessment guidelines were positively correlated to a considerable and statistically significant degree. Unsurprisingly, scores on the two variants of the Static-99 were also positively and significantly correlated. Correlations between the SPJ guidelines and actuarial tools, while statistically significant, were not of the same magnitude.

As a measure of internal consistency, the relationship between the total scores on the two SPJ risk assessment guidelines and their structured risk ratings/case prioritisation levels was investigated. A monotonic relationship between the number of risk factors present and the final risk judgements was expected. SVR-20 total scores were significantly and positively correlated with the overall risk judgement ($\tau_b = .49$, $p < .01$). For the RSVP, lifetime scores were significantly and positively correlated with case prioritisation judgements ($\tau_b = .39$, $p < .01$). Case prioritisation judgements were significantly correlated to a similar degree with both the summed score of past item ratings on the RSVP ($\tau_b = .37$, $p < .01$) and current item ratings ($\tau_b = .45$, $p < .01$).

Table 4 displays correlations between the five domains of the RSVP and the four-domain conceptualisation of the SVR-20.

While the factors that form each domain differ between the RSVP and SVR-20, it can be seen that there were strong correlations between the domains. The strongest relationships make theoretical sense as the factors contained within the domains are similar. The RSVP's Mental Disorder domain is most related to the SVR-20's Psychological Adjustment domain as the factors contained in these domains are similar, e.g. sexual deviance, major mental illness, substance abuse problems and suicidal ideation. The Sexual Offending domains of the two instruments unsurprisingly correlate highly with overlapping factors such as chronicity, diversity and physical harm. The overlapping factors that account for the high correlations in the Social Adjustment domains include relationship factors, and the RSVP's Manageability domain is highly correlated with the SVR-20's Future Plans domain in relation to planning and attitudes towards treatment. The lower correlation between the Psychological Adjustment domains reflects differences between the domain factors, as the items in that domain on the SVR-20 are more located in the Mental Disorder domain of the RSVP such as sexual deviation, major mental illness and substance use. It is clear that the two instruments are best considered equivalent or parallel forms (Hart & Boer, 2009).

Table 3. Pearson's correlations (r) between total scores on the different tools.

	SVR-20 Total Score	RSVP Lifetime Total Score	Static-99 Total Score	Static-99R Total Score
SVR-20	-			
RSVP	.93**	-		
Static-99	.41**	.37**	-	
Static-99R	.38**	.32**	.91**	-

Note. $N = 95$. * = $p < .05$; ** $p < 0.01$.

Table 4. Pearson's correlations (*r*) between domains of assessment tools.

	SVR-20 Psychological Adjustment	SVR-20 Social Adjustment	SVR-20 Sexual Offences	SVR-20 Future Plans
RSVP Sexual Violence History	.39**	.24*	.70**	.16
RSVP Psychological Adjustment	.48**	.19	.64**	.32**
RSVP Mental Disorder	.80**	.39**	.22*	.36**
RSVP Social Adjustment	.39**	.85**	.22*	.60**
RSVP Manageability	.31**	.72**	.23*	.93**

Note: * = $p < .05$; ** $p < 0.01$.

Predictive validity

Case prioritisation and risk judgements

ROC analyses found that neither the RSVP case prioritisation rating ($AUC = .57$) nor the SVR-20 overall judgement ($AUC = .57$) were able to differentiate between recidivists and non-recidivists in a statistically significant manner. Given the identical effect size, unsurprisingly there was no statistically significant differences between the AUC values obtained for the two tools ($z = .03$, $p = .98$; DeLong et al., 1988). Both the Static-99 ($AUC = .76$) and the Static-99R ($AUC = .77$) performed more strongly in discriminating between recidivists and non-recidivists. Confidence intervals were wide for the RSVP case prioritisation and SVR-20 overall judgement and several low scores were found with both tools during this analysis. It was also found that the RSVP case prioritisation ratings and the SVR-20 risk judgements had similar predictive validity at five years but improved at 10 years. These results are included in Table 5. It can be seen that both judgements provided better discrimination at the 10-year mark, though still non-significant. It is also noted that the RSVP reduces at 15 years to the same value as at the five-year mark.

For easy comparison with Vargen et al. (2020) relative recidivism rates were also calculated for each of the RSVP risk categories described in Table 1. Using the low risk group as a reference the increased odds of recidivism (as indexed using odds ratios) were 1.80 times greater in the high risk category, but only 0.97 times in the moderate risk category. The overall difference in recidivism rates between the three groups was not statistically significant ($\chi^2(2, N = 95) = 1.08$, $p = .58$).

Differences in time to sexual reoffending were analysed using Kaplan-Meier survival analysis with log-rank tests to assess whether there was a difference between each of the risk/priority categories over time. While the proportion of recidivists in all categories

Table 5. AUC Values for RSVP Case Prioritisation and SVR-20 Risk Judgements After 5, 10, and 15 Years at Risk.

	5 Years n = 83	10 Years n = 62	15 Years n = 29
RSVP			
AUC	.54	.60	.54
SE	.09	.09	.11
95%CI	[.36, .70]	[.43, .76]	[.33, .76]
SVR-20			
AUC	.53	.61	.60
SE	.09	.09	.11
95%CI	[.36, .70]	[.45, .78]	[.39, .81]

Note: AUC = Area Under the Curve (95% confidence intervals are asymptotic).

increased over time, and was greatest in the high risk category, there were no significant differences in time to reoffending between RSVP case prioritisation categories ($\chi^2(2, N = 95) = 1.3, p = .5$) or between SVR-20 risk judgement categories ($\chi^2(2, N = 95) = 1.18, p = .6$).

Total and domain scores

While the case prioritisation (RSVP) and risk judgements (SVR-20) did not have a statistically significant relationship with recidivism, it was also important to investigate whether the total 'scores' on the instruments, and their domains, were related to sexual reoffending. Indeed, the literature regarding SPJ guidelines has routinely investigated the mechanical use of the various tools in addition to structured judgements as SPJ guidelines are more susceptible to individual rater effects.

The relationship between RSVP scores in each domain, and in each temporal coding period (i.e. past, recent, and lifetime), and sexual recidivism is presented in Table 6. Also presented are the results for the SVR-20 total scores, the change score, and a composite measure that was calculated by adding the total score with the change score (e.g. a total score of X and a change score of Y are added to obtain a composite score of Z). The results of the RSVP and SVR-20 scores are also provided at five and 10-year follow-up periods, along with the domain scores for both instruments.

It can be seen that the RSVP total scores were better able to discriminate between recidivists and non-recidivists than the structured judgements. However, only the recent total score had a statistically significant relationship with sexual recidivism for the RSVP, both over the total follow-up period and at 10 years. Moreover, the *Manageability* domain was the only RSVP domain that was consistently and significantly related to recidivism. Similarly, the SVR-20 total score had a moderate relationship with sexual recidivism, reaching statistical significance at the 10-year follow-up period. The SVR-20 composite score varied between having a limited to no relationship with recidivism. The *Future Plans* domain of the SVR-20 was significantly related to recidivism.

Table 6. Ability of RSVP and SVR-20 total and domain scores to discriminate between recidivists and non-recidivists.

RSVP	Past		Recent		Lifetime	
	AUC	95%CI	AUC	95%CI	AUC	95%CI
Total follow-up	59 [.44, .74]		65 [.51, .79]		63 [.48, .77]	
5 Years	52 [.36, .68]		61 [.46, .76]		56 [.40, .71]	
10 Years	61 [.45, .77]		67 [.53, .82]		64 [.49, .79]	
Sexual Violence History	51 [.36, .67]		60 [.46, .75]		56 [.41, .71]	
Psychological Adjustment	59 [.46, .73]		62 [.49, .76]		60 [.48, .73]	
Mental Disorder	53 [.37, .69]		58 [.44, .73]		55 [.40, .70]	
Social Adjustment	55 [.38, .73]		57 [.40, .75]		57 [.39, .74]	
Manageability	66 [.52, .81]		70 [.55, .84]		67 [.52, .82]	
SVR-20	Total Score		Recent Change		Total Score and Recent Change	
	AUC	95%CI	AUC	95%CI	AUC	95%CI
Total Scores	65 [.50, .79].		55 [.41, .69]		47 [.31, .64]	
5 Years	60 [.45, .76].		60 [.45, .76]		55 [.37, .73]	
10 Years	68 [.53, .83].		56 [.40, .72]		56 [.39, .73]	
Psychological Adjustment	58 [.44, .71].		50 [.34, .67]		57 [.44, .70]	
Social Adjustment	56 [.39, .73].		50 [.34, .65]		55 [.38, .72]	
Sexual Offending	58 [.43, .73].		62 [.48, .77]		61 [.46, .75]	
Future Plans	72 [.58, .87].		54 [.39, .69]		73 [.59, .86]	

Note: AUC = Area Under the Curve (95% confidence intervals are asymptotic). Total $N = 95$. 5 year $n = 83$, 10 year $n = 62$.

What risk factors may be most predictively useful with older men?

The majority of the RSVP risk factors were correlated to a statistically significant degree with the rater's case prioritisation level, but few factors or domains were independently correlated with sexual recidivism. As noted above, the *Manageability* domain was statistically significant ($r_{pb} = .27, p < .01$) and contained within that domain, it was the factor *Problems with Planning* that was particularly correlated with recidivism ($r_{bp} = .24, p < .05$). The only other individual factor that was correlated with recidivism was *Problems with Stress and Coping* ($r_{pb} = .24, p < .05$). The relationships with sexual recidivism for these factors were of such a magnitude that potentially they may provide some useful indications of risk factors when assessing the risk of older sexual offenders. To this end, it was important to identify whether the items captured by the SVR-20 *Future Plans* domain (lacks realistic plans, negative attitude toward intervention) and the RSVP item *Problems with stress or coping* were not simply accounting for the same variance as the Static-99R. We found that a summing of these item scores was significantly correlated with the Static-99R score ($r = .38, p < .001$). However, partial correlations showed that the two scores accounted for unique proportions of the variance in sexual recidivism. The Static-99R correlation ($r_p = .39, p < .001$) reduced slightly when these three items were controlled for, but were still statistically significant ($r_p = .29, p < .001$). Similarly, the correlation of the three items with recidivism ($r_p = .36, p < .001$) reduced slightly when the Static-99R was controlled for, but also remained significant ($r_p = .24, p < .05$). Accordingly, these future-oriented dynamic risk factors, at least in this sample, were accounting for unique variance that was not encompassed by the Static-99R.

Discussion

The aims of the current study were to evaluate the inter-rater reliability, internal consistency, convergent and predictive validity of the RSVP and SVR-20 guidelines in an Australian population of older men who had offended sexually. The aim of convergent validity for these two SPJ risk assessment guidelines was very strong in regard to scores, domain totals, and structured judgements; and supported the contention that they are indeed equivalent or parallel forms of structured risk guidelines (Hart & Boer, 2009). Both had similar levels of inter-rater reliability, which was consistent with previous studies with wider age ranges (e.g. Hart & Boer, 2009; Kanters et al., 2017; Sutherland et al., 2012; Vargen et al., 2020). The risk factors contained within both tools were strongly related to assessor's risk judgements. However, the assessor's judgements made with either guideline had very limited ability to discriminate between recidivists and non-recidivists over the entire follow-up period, with noted differences observed for the RSVP as the results for five and 15 years were the same, while the SVR-20 did not reflect similarities between the three follow-up periods. The reasons for this finding are not immediately clear.

The examination of inter-rater reliability results as per the aims of the study, indicated that the scoring of the SPJ guidelines risk assessment was not influenced by user bias or error and was consistent with previous assessors who actually interviewed the participants. There was also clear evidence of internal consistency, with case prioritisation/judgements being correlated with total scores (Boer & Hart, 2009; Kanters et al., 2017). This

suggests the failure of the structured judgements was not likely due to anomalies of judgement or scoring. When used to guide structured judgements about future risk, SPJ risk assessment guidelines did not perform as well in the current older sample as for the younger sample described in Davis et al. (in submission) study of the RSVP. They also did not perform as well as the actuarial Static-99 or Static-99R (see Raymond et al., 2020).

The current study found that there was some improvement in predictive ability when using lifetime total scores (RSVP AUC = .63, SVR-20 AUC = .65), which was particularly clear at the 10-year follow-up point. This was consistent with the study by Davis et al. (2024; RSVP AUC = .73) who used an adult sample drawn from the same location and assessment period as the current study. Notably, for the RSVP, it was scores derived from the 'recent' ratings of risk factors that were more strongly related to sexual recidivism, even over a decade of follow-up, however, at 15 years the scores fell to the same level as at five years. Given these scores are meant to reflect more dynamic (changeable) risk, their long-term predictive capacity was unexpected and the reasons for this finding not immediately clear. It is possible that these scores better capture the older men's current situation (including age-related changes), which is more closely related to their future situation in which subsequent offending occurs, than past scores that reflect the totality of their life up to the point of assessment.

Review of the relationship between domain and total scores, and sexual recidivism, showed the *Manageability/Future Plans* domains were the only domains consistently discriminating between older recidivists and non-recidivists in both sets of guidelines (AUC = .67 and .75), with significant correlations between domain scores and sexual recidivism. Interestingly, in Davis et al. (in submission) study of the performance of the RSVP guidelines in men under 50, the *Mental Disorder* and *Manageability* domains both discriminated between recidivists and non-recidivists, with the sexual deviance item in the *Mental Disorder* domain being a particularly potent risk factor. The differences between the studies of the two age groups indicated that while the RSVP targeted adult risk factors, it was less robust with the older sample. Davis et al. did not examine the performance of the SVR-20 in the younger sample. While admittedly speculative, the fact that the current sample of older offenders was primarily comprised of those with child victims may have contributed to the sexual deviance item, and the *Mental Disorder* domain, being largely unrelated to recidivism in the current study.

Examination of the relationship between domain and risk factor scores and recidivism suggested some possible directions for future research with older men who sexually offend. In particular, the RSVP factor *Problems with Stress and Coping*, and SVR-20 Future Plans domain (consisting of *Lacks Realistic Plans* and *Negative Attitude Toward Intervention*) were independently related to sexual recidivism in this sample and accounted for a unique portion of the variance when compared to the predictive power of the Static-99R. While this is a result that demands replication, the magnitude of relationships suggests that these may be useful dynamic risk factors to concentrate on for both future research and applied forensic practice with older sexual offenders. Given that the current study is one of the few in the literature to address the validity of risk factors with an older population of men who have sexually offended, these admittedly preliminary results provide a fertile ground for further study.

Limitations/Implications

While this study contributes to the wider literature about the utility of the RSVP and SVR-20, it has some important limitations that must be borne in mind. There are several limitations associated with the sample. The small sample size, albeit large in comparison to other research, restricts generalising conclusions to the general population of older men who have sexually offended. The file information was sourced from a specialist forensic community centre and may be a limitation. Specifically, at the time of offender attendance (between 1987 and 2011), the clinic was one of the only referral services that provided assessment and treatment to those who had offended sexually and as a result, were more likely to be higher risk offenders.

Research into the validity of decisions made using SPJ guidelines, specifically with larger samples, will produce results that can be more readily generalised. Related to the small sample size is the fact that we were unable to separately investigate whether the guidelines performed differently for men who had offended historically versus men whose first sexual offences occurred in older age. There are reasons to hypothesise that sexual offending trajectories may differ for these two groups, but this could not be investigated in this study. A further potential limitation related to the sample is that it was overwhelmingly composed of men who had offended sexually against children. This is consistent with past research on older men who have sexually offended (Rodriguez et al., 2017), but may limit the generalisability of the results. Future research focusing on victimology would be beneficial as victim selection is not always associated with deviance but related to psychosocial factors and access (Miller, 2013; Nicholaichuk et al., 2014) which could impact on assessment of risk and management for older offenders.

The second source of limitations was the file review methodology, leading to a paucity of complex and sophisticated information in some files, particularly those relating to assessments conducted in the early 1990s. Hart and Boer (2009), who were involved in the development of the SVR-20 and RSVP, have criticised previous research with regard to the use of data obtained from files versus field research. The file review method also meant we were unable to rate the psychopathy risk factor from the SVR-20 and the RSVP. Darjee and colleagues (2016) found that while the psychopathy factor is not solely responsible for the predictive validity of the RSVP, it did play an important role, meaning its presence may have produced different results in the current study. It must be stressed that the file review methodology of the current research also meant that we were unable to thoroughly investigate how the full structured process of developing conclusory opinions using the RSVP performed. The current study focused on evaluating the inter-rater reliability, convergent validity, and predictive validity of the SVR-20 and RSVP items and structured risk judgements. In this, it was essentially a study of the 'lite' use of SPJ guidelines (Logan & Lloyd, 2019). Accordingly, we were unable to examine how ratings of future relevance or future risk scenarios related to recidivism. These are identified by the authors as being important to the predictive performance of the RSVP (though are not part of the SVR-20). Given some of the results suggested the possible importance of contextual and dynamic risk factors, these aspects of the SPJ model could be particularly important when applying the RSVP and related schemes to older sexual offenders.

Conclusion

The current study adds to the relatively limited literature on the validity of the SVR-20 and RSVP, and is the first to assess these tools specifically in a sample of older men with histories of sexual offending. The results addressed the aims of the study related to interrater reliability, internal consistency, convergent and predictive validity. The results also indicated that risk/case prioritisation judgements arising from these structured guidelines did not perform well in discriminating between those with and without sexual recidivism. However, the results do suggest that the total scores of risk factors on the SVR-20, and the recent status of risk factors on the RSVP, may have a moderate relationship to long-term sexual recidivism. Moreover, specific dynamic items from these guidelines relating to future planning and engagement with intervention do appear to have a predictive relationship with sexual recidivism in this group, and account for unique variance not accounted for by the Static-99R. While all of these results are admittedly tentative given the sample size and the file-review methodology, they both demand replication whilst still offering some fruitful directions for further research and applied forensic practice.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Australian Institute of Health and Welfare. (2020). *Health and ageing of Australia's prisoners 2018*. Cat no PHE269. <https://www.aihw.gov.au/reports/prisoners/health-and-ageing-of-australias-prisoners-2018>.
- Baidawi, S., Turner, S., Trotter, C., Browning, C., Collier, P., O'Connor, D., & Sheehan, R. (2011). Older prisoners: A challenge for Australian corrections. In *Trends & issues in crime and criminal justice*. No. 426. Australian Institute of Criminology. <https://www.aic.gov.au/publications/tandi/tandi426>.
- Barbaree, H. E., Langton, C. M., Blanchard, R., & Boer, D. P. (2008). Predicting recidivism in sex offenders using the SVR-20: The contribution of age-at-release. *International Journal of Forensic Mental Health*, 7(1), 47–64. <https://doi.org/10.1080/14999013.2008.9914403>
- Bauer, M., Nay, R., & McAuliffe, L. (2009). Catering to love, sex and intimacy in residential aged care: What information is provided to consumers? *Sexuality and Disability*, 27(1), 3–9. <https://doi.org/10.1007/s11195-008-9106-8>
- Beech, A. R., Wakeling, H. C., Szumski, F., & Freemantle, N. (2016). Problems in the measurement of dynamic risk factors in sexual offenders. *Psychology, Crime & Law*, 22(1-2), 68–83. <https://doi.org/10.1080/1068316X.2015.1109095>
- Bengtson, S., Lund, J., Ibsen, M., & Langstrom, N. (2019). Long-term violent reoffending following forensic psychiatric treatment: Comparing forensic psychiatric examinees and general offender controls. *Frontiers in Psychiatry*, 10(715). <https://doi.org/10.3389/fpsy.2019.00715>.
- Blackley, R., & Bartels, L. (2018). Sentencing and treatment of juvenile sex offenders in Australia (Report No. 555). In *Trends & issues in crime and criminal justice*. Australian Institute of Criminology. <https://www.aihw.gov.au/reports/prisoners/health-and-ageing-of-australias-prisoners-2018>.
- Boer, D. (2006). Sexual offender risk assessment strategies: Is there a convergence of opinion yet? *Sexual Offender Treatment*, 1(2), 1–4.
- Boer, D. P., & Hart, S. D. (2009). Sex offender risk assessment: Research, evaluation, 'best-practice' recommendations and future directions. In J. L. Ireland, C. A. Ireland, & P. Birch (Eds.), *Violent and sexual offenders: Assessment, treatment and management* (pp. 27–42). Routledge.

- Boer, D., Hart, S., Kropp, P., & Webster, C. (1997). *Manual for the sexual violence risk-20*. Burnaby, BC: The British Columbia Institute Against Family Violence, Co-published with the Mental Health, Law, and Policy Institute at Simon Fraser University.
- Boer, D. P., Tough, S., & Haaven, J. (2004). Assessment of risk manageability of intellectually disabled sex offenders. *Journal of Applied Research in Intellectual Disabilities*, 17(4), 275–283. <https://doi.org/10.1111/j.1468-3148.2004.00214.x>
- Cheng, J., Haag, A., & Olver, M. (2019). Predictors of Historical Clinical Risk Management-20 Version 3 (HCR-20:V3) summary risk ratings. *Psychiatry, Psychology and Law*, 26(1), 682–692. <https://doi.org/10.1080/13218719.2019.1618753>
- Chua, J., Gheung, G., Hatters Friedman, S., & Taylor, S. (2018). What are the characteristics of first time older sexual offenders? *International Psychogeriatrics*, 30(5), 647–662. <https://doi.org/10.1017/S1041610217002186>
- Cortoni, F., Babchishin, K. M., & Rat, C. (2017). The proportion of sexual offenders who are female is higher than thought: A meta-analysis. *Criminal Justice and Behavior*, 44(2), 145–162. <https://doi.org/10.1177/0093854816658923>
- Cortoni, F., Hanson, R. K., & Coache, M. (2010). The recidivism rates of female sexual offenders are low: A meta-analysis. *Sexual Abuse: A Journal of Research and Treatment*, 22(4), 387–401. <https://doi.org/10.1177/1079063210372142>
- Craig, L., Browne, K. D., & Stringer, I. (2004). Comparing sex offender risk assessment measures on a UK sample. *International Journal of Offender Therapy and Comparative Criminology*, 48(1), 7–27. <https://doi.org/10.1177/0306624X03257243>
- Darjee, R., Russell, K., Forrest, L., Milton, E., Savoie, V., Baron, E., Kirkland, J., & Stobie, S. (2016). Risk for Sexual Violence Protocol (RSVP): A real world study of the reliability, validity and utility of a structured professional judgement instrument in the assessment and management of sexual offenders in South East Scotland. Risk Management Authority and NHS Scotland Forensic Network. <https://www.rma.scot/risk-of-sexual-violence-protocol-rsvp-study-published/>
- Davis, M. R., Woods, M., Raymond, B., Ogloff, J. R. P., & McEwan, T. E. (in submission). Does it work? Convergent and PREDICTIVE VALIDITY OF the Risk for Sexual Violence Protocol (RSVP) in an Australian Community Forensic Mental Health Setting. Manuscript submitted for publication.
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44(3), 837–845. <https://doi.org/10.2307/2531595>
- 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. <https://doi.org/10.1023/B:LAHU.0000029137.41974.eb>
- Fedoroff, J. P., Richards, D., Ranger, R., & Curry, S. (2016). The predictive validity of common risk assessment tools in men with intellectual disabilities and problematic behaviors. *Research in Developmental Disabilities*, 57, 29–38. <https://doi.org/10.1016/j.ridd.2016.06.011>
- Flieiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, 5(1), 105–112. <https://doi.org/10.1177/014662168100500115>
- Gunby, C., & Woodhams, J. (2010). Sexually deviant juveniles: Comparisons between the offender and offence characteristics of ‘child abusers’ and ‘peer abusers’. *Psychology, Crime & Law*, 16(1–2), 47–64. <https://doi.org/10.1080/10683160802621966>
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>
- Hanson, R. K. (2001). *Age and sexual recidivism: A comparison of rapists and child molesters*. Report No. 2001-01. Solicitor General Canada.
- Hanson, R. K. (2002). Recidivism and age: Follow-up data from 4,673 sexual offenders. *Journal of Interpersonal Violence*, 17(10), 1046–1062. <https://doi.org/10.1177/088626002236659>
- Hanson, R. K. (2006). Does Static-99 predict recidivism among older sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 18(4), 343–355. <https://doi.org/10.1007/s11194-006-9027-y>

- 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. <https://doi.org/10.1037/a0014421>
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24(1), 119–136. <https://doi.org/10.1023/A:1005482921333>
- Hare, R. D. (2003). *The Hare Psychopathy Checklist – revised* (2nd ed.). Multi-Health Systems.
- Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. (2003). Static-99 coding rules revised-2003. https://amyphenix.com/wp-content/uploads/2016/02/Static-99_Coding_Rules.pdf
- Harris, A., Rawson, J., & Hanson, R. K. (2010). Clinical, actuarial and dynamic risk assessment of sexual offenders: Why do things keep changing? *Journal of Sexual Aggression*, 16(3), 296–310. <https://doi.org/10.1080/13552600.2010.494772>
- Harris, G. T., Rice, M. E., Quinsey, V. L., & Cormier, C. A. (2015). The actuarial prediction of violence. In *Violent offenders: Appraising and managing risk* (3rd ed., pp. 121–168). American Psychological Association. <https://doi.org/10.1037/14572-000>
- Hart, S. D., & Boer, D. P. (2009). 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). Routledge.
- Hart, S., Kropp, P., Laws, D., Klaver, J., Logan, C., & Watt, K. (2003). *The Risk for Sexual Violence Protocol (RSVP): Structured professional guidelines for assessing risk of sexual violence*. Mental Health, Law, and Policy Institute at Simon Fraser University.
- Harris, G. T., & Rice, M. E. (2007). Adjusting actuarial violence risk assessments based on aging or the passage of time. *Criminal Justice and Behavior*, 34(3), 297–313. <https://doi.org/10.1177/009385486293486>
- 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 weight. *Sexual Abuse: A Journal of Research and Treatment*, 24, 64–101. <https://doi.org/10.1177/1079063211409951>
- Holoyda, B. J., & Newman, W. J. (2016). Recidivism risk assessment for adult sexual offenders. *Current Psychiatry Reports*, 18(17), 1–7. <https://doi.org/10.1007/s11920-015-0650-5>
- Kanters, T., Homsveld, R. H. J., Nunes, K. L., Zwets, A. J., Muris, P., & van Marle, H. J. C. (2017). The Sexual Violence Risk-20: Factor structure and psychometric properties. *The Journal of Forensic Psychiatry & Psychology*, 28(3), 368–387. <https://doi.org/10.1080/14789949.2017.1284887>
- Kelley, S. M., Ambroziak, G., Thornton, D., & Barahal, R. M. (2018). How do professionals assess sexual recidivism risk? An updated survey of practices. *Sexual Abuse*, 32(1), 3–29. <https://doi.org/10.1177/1079063218800474>
- Kelley, S. M., & Thornton, D. (2015). Assessing risk of sex offenders with major mental illness: Integrating research into best practices. *Journal of Aggression, Conflict and Peace Research*, 7(4), 258–274. <https://doi.org/10.1108/JACPR-02-2015-0162>
- Landis, J. R., & Koch, G. G. (1977). An application of hierarchical Kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 33(2), 363–374. <https://doi.org/10.2307/2529786>
- Lindsay, W. R., & Taylor, J. L. (2010). Understanding and treating offenders with learning disabilities: A review of recent developments. *Journal of Learning Disabilities and Offending Behaviour*, 1(1), 5–16. <https://doi.org/10.5042/jldob.2010.0173>
- Logan, C., & Lloyd, M. (2019). Violent extremism: A comparison of approaches to assessing and managing risk. *Legal and Criminological Psychology*, 24(1), 141–161. <https://doi.org/10.1111/lcrp.12140>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia medica*, 22(3), 276–282. <https://doi.org/10.11613/BM.2012.031>
- Miller, L. (2013). Sexual offenses against children: Patterns and motives. *Aggression and Violent Behavior*, 18(2013), 506–519. <https://doi.org/10.1016/j.avb.2013.07.006>
- Nicholaichuk, T. P., Olver, M., Gu, D., & Wong, S. C. (2014). Age, actuarial risk, and long-term recidivism in a national sample of sex offenders. *Sexual Abuse*, 26(5), 406–428. <https://doi.org/10.1177/1079063213492340>

- Ogloff, J. R. P. & Davis, M. R. (2020). From predicting dangerousness to assessing and managing risk for violence. In Wormith, S., Craig, L. A., & Hogue, T. E. (Eds.) *The Wiley Handbook of what works in violence risk management: Theory, research and practice* (pp. 81–98). John Wiley and Sons. <https://doi.org/10.1002/9781119315933>
- Olver, M. E., Beggs Christofferson, S. M., Nicholaichuk, T. P., & Wong, S. C. P. (2020). Predictive properties of the Violence Risk Scale–Sexual Offence version as a function of age. *Assessment*, 28(6), 1671–1693. <https://doi.org/10.1177/10731911020914405>
- Olver, M. E., Nicholaichuk, T. P., Gu, D., & Wong, S. C. P. (2012). Sex offender treatment outcome, actuarial risk, and the aging sex offender in Canadian Corrections: A long-term follow-up. *Sexual Abuse*, 25(4), 396–422. <https://doi.org/10.1177/1079063212464399>
- Ozkan, T., Clipper, S. J., Piquero, A. R., Baglivio, M., & Wolff, K. (2020). Predicting sexual recidivism. *Sexual Abuse*, 32(4), 375–399. <https://doi.org/10.1177/1079063219852944>
- Phenix, A., Fernandez, Y., Harris, A. J. R., Helmus, M., Hanson, R. K., & Thornton, D. (2016). Static-99R coding rules revised – 2016. http://static99.org/pdfdocs/Coding_manual_2016
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. (1998). *Violent offenders: Appraising and managing risk*. American Psychological Association.
- Rakes, S., Prost, S. G., & Tripodi, S. J. (2018). Recidivism among older adults: Correlates of prison re-entry. *Justice Policy Journal*, 15(1), 1–15.
- Ralston, C. A., & Epperson, D. L. (2013). Predictive validity of adult risk assessment tools with juveniles who offended sexually. *Psychological Assessment*, 25(3), 905–916. <https://doi.org/10.1037/a0032683>
- Raymond, B. C., McEwan, T. E., Davis, M. R., Reeves, S., & Ogloff, J. R. P. (2020). Investigating the predictive validity of the Static-99/99R with older sexual offenders. *Psychiatry, Psychology and Law*, 28(1), 120–134. <https://doi.org/10.1080/13218719.2020.1767714>
- Reeves, S. G., Ogloff, J. R. P., & Simmons, M. (2017). The predictive validity of the Static-99, Static-99R, and Static-2002/R: Which one to use? *Sexual Abuse*, 30(8), 887–907. <https://doi.org/10.1177/1079063217712216>
- 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. <https://doi.org/10.1177/0093854811416908>
- Rettenberger, M., Haubner-Maclean, T., & Eher, R. (2013). The contribution of age to the Static-99 risk assessment in a population-based prison sample of sexual offenders. *Criminal Justice and Behavior*, 40(12), 1413–1433. <https://doi.org/10.1177/0093854813492518>
- Rettenberger, M., & Hucker, S. J. (2011). Structured professional guidelines: International applications. In D. P. Boer, R. Eher, L. A. Craig, M. H. Miner, & F. Pfafflin (Eds.), *International perspectives on the assessment and treatment of sexual offenders: Theory, practice, and research* (pp. 269–294). John Wiley & Sons.
- Rettenberger, M., Hucker, S. J., Boer, D. P., & Eher, R. (2009a). The reliability and validity of the Sexual Violence Risk-20 (SVR-20): An international review. *Sexual Offender Treatment*, 4(2), 1–14.
- Rettenberger, M., Matthews, A., Boer, D. P., & Eher, R. (2009b). 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. <https://doi.org/10.1177/0306624X08328755>
- Rodriguez, M., Boyce, P., & Hodges, J. (2017). A neuropsychological study of older adult first-time sex offenders. *Neurocase*, 23(2), 154–161. <https://doi.org/10.1080/13554794.2017.1334802>
- Sea, J., & Hart, S. D. (2021). Interrater reliability and concurrent validity of the Risk for Sexual Violence Protocol for Korean sexual offenders: A field study.
- Shaw, R., Stevens, B., Paget, J., & Snoyman, P. (2018). Ageing in corrective services: From the perspective of prison chaplains. *Psychiatry, Psychology and Law*, 26(1), 97–109. <https://doi.org/10.1080/13218719.2018.1483275>
- Singh, J. P., Yang, S., Mulvey, E. P., & the RAGEE Group. (2015). Reporting guidance for violence risk assessment predictive validity studies: The RAGEE statement. *Law and Human Behavior*, 39(1), 15–22. <https://doi.org/10.1037/lhb0000090>

- Stevens, P., Hutchin, K., French, L., & Craissati, J. (2013). Developmental and offence-related characteristics of different types of adolescent sex offender: A community sample. *Journal of Sexual Aggression, 19*(2), 138–157. <https://doi.org/10.1080/13552600.2011.645889>
- Sutherland, A. A., Johnstone, L., Davidson, K. M., Hart, S. D., Cooke, D. J., & Kropp, P. (2012). Sexual violence risk assessment: An investigation of the interrater reliability of professional judgments made using the Risk of Sexual Violence Protocol. *International Journal of Forensic Mental Health, 11*(2), 119–133. <https://doi.org/10.1080/14999013.2012.690020>
- Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). Distinguishing and combining risks for sexual and violent recidivism. *Annals of the New York Academy of Sciences, 989*(1), 225–235. <https://doi.org/10.1111/j.1749-6632.2003.tb07308.x>
- Vargen, L. M., Jackson, K. J., & Hart, S. D. (2020). Interrater reliability, concurrent validity, and predictive validity of the Risk for Sexual Violence Protocol. *Law and Human Behavior, 44*(1), 37–50. <https://doi.org/10.1037/lhb0000353>
- Wong, S., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. (2017). *The violence risk scale: Sexual offense version (VRS-SO)*. Saskatoon, Saskatchewan, Canada: Regional Psychiatric Centre and University of Saskatchewan.
- Worling, J. R. (2004). The Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR): Preliminary psychometric data. *Sexual Abuse, 16*(3), 235–254. <https://doi.org/10.1177/107906320401600305>
- Worling, J. R., Bookalam, D., & Litteljohn, A. (2012). Prospective validity of the Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR). *Sexual Abuse, 24*(3), 203–223. <https://doi.org/10.1177/1079063211407080>
- Worling, J. R., & Curwen, T. (2000). Adolescent sexual offender recidivism: Success of specialized treatment and implications for risk prediction. *Child Abuse & Neglect, 24*(7), 965–982. [https://doi.org/10.1016/S0145-2134\(00\)00147-2](https://doi.org/10.1016/S0145-2134(00)00147-2)
- Youngstrom, E. A. (2013). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology, 39*(2), 204–221. <https://doi.org/10.1093/jpepsy/jst062>