Predictors of Historical Clinical Risk Management-20 Version 3 (HCR–20:V3) summary risk ratings

Jeremy Cheng^a, Andrew M. Haag^{b,c,d} and Mark E. Olver^a

^aDepartment of Psychology, University of Saskatchewan, Saskatoon, SK, Canada; ^bAlberta Health Services, Edmonton, AB, Canada; ^cDepartment of Psychiatry, University of Alberta, Edmonton, AB, Canada; ^dDepartment of Psychology, University of Alberta, Edmonton, AB, Canada

How individual risk factors on structured professional judgement (SPJ) assessment tools translate into SPJ final risk formulations is unclear due to a lack of structured criteria. Understanding pathways to risk formulations is vital, as they serve as intervention targets for risk management. This study examined how Historical Clinical Risk Management-20 Version 3 (HCR–20:V3) raters weighed varied information sources to complete summary risk ratings (SRRs). Four independent raters retrospectively coded an archived sample of 32 inpatients at a Canadian forensic psychiatric hospital. HCR–20:V3 SPJ SRRs were regressed on the 20 individual items and sample covariates to identify unique predictors of risk formulations across each rater. Raters consistently used HCR–20:V3 items and composite subscales for SRRs. Despite strong inter-rater agreement on the SRRs, there were variations across raters regarding which items informed each SRR. Rater-unique biases were also shown to influence SRRs. Implications for forensic practice and risk management are discussed.

KEYWORDS: HCR–20:V3; risk assessment; summary risk ratings; structured professional judgement; risk management; final risk formulations.

Violence risk assessment is a prominent component of forensic mental health and criminal justice practice, as courts rely on risk evaluators to determine facts on dangerousness and public safety (Borum, 1996; Latimer & Lawrence, 2006). Over the past two decades, the field of risk assessment has expanded its focus from the prediction of recidivism to greater emphasis on risk management (Müller-Isberner, Webster, & Gretenkord, 2007) – that is, how risk may be mitigated to prevent violent victimization. As this trend towards risk management has grown, dynamic risk assessment tools have been developed to assist individualized intervention plans to reduce risk (Logan, 2014).

Risk management is particularly meaningful in decision-making contexts where curtailing public safety risks take priority over other factors in community treatment such as physical health (Mossman, 2000). Indeed, risk management has been described as an ethical, moral, and legal imperative for forensic clinicians (Hart, 1998). Moreover, from an economic perspective, high-risk individuals tend to utilize costly services in the health care system such as inpatient hospitalization (Borum, 1996). It follows that risk reduction should in principle lead to cost savings from shorter service use and by effective reintegration of men and women into society.

Correspondence: Andrew M. Haag, Alberta Health Services, Department of Psychiatry and Department of Psychology, 17480 Fort Road, Edmonton, AB T5J 2J7, Canada. E-mail: haag1@ualberta.ca

Structured professional judgement (SPJ) measures are a tradition of risk assessment tools that integrate empirically supported risk factors with clinical judgement to generate a final risk judgement for the potential of recidivism outcomes such as violence. They are sometimes contrasted with actuarial risk assessment instruments (ARAIs), which yield numeric scale ratings from summing the risk items, which in turn are linked to recidivism estimates over a defined follow-up period; they may be static, as in the case of the Violence Risk Appraisal Guide-Revised (VRAG-R; Harris, Rice, Quinsey, & Cormier, 2015), or dynamic and capable of changing, such as the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995).

The HCR-20:V3 is a SPJ measure designed to assess risk for future violence and to inform risk management planning to prevent future violence. It is composed of 10 static (historical) and 10 dynamic (clinical and risk) items and has applications in institutional (i.e. correctional, forensic, and general or civil psychiatric) and community settings (Douglas et al., 2014). Extensive research supports the psychometric properties of the HCR-20:V3 and its previous versions, including inter-rater reliability (Bjorkly, Eidhammer, & Selmer, 2014; de Vogel, van den Broek, & de Vries Robbé, 2014; Douglas & Belfrage, 2014; Doyle et al., 2014; Howe, Rosenfeld, Foellmi, Stern, & Rotter, 2016; Kötter et al., 2014; Smith, Kelley, Rulseh, Sörman, & Edens, 2014) and predictive validity for future violence (Douglas, 2014; Doyle et al., 2014).

There are several important clinical features of the HCR-20:V3 (Farrington, Jolliffe, & Johnstone, 2008), including its ability to track treatment change, focus on risk management to reduce future risk, and emphasize clinical expertise (Logan, 2014), although only a small collection of studies have heretofore examined the dynamism of the tool. The tool provides trichotomous summary risk ratings (SRRs) using the traditional categories of low, moderate and high, which represent a risk

formulation rather than a probabilistic estimate of future violence risk. That is, the HCR-20:V3 utilizes clinical judgement to synthesize item risk ratings into SRRs for scenario planning around future violence, severity, imminence and the likelihood of incident (Douglas et al., 2014). The function of SRRs are to alert the clinician to specific harmful outcomes so that risk management strategies can tailor interventions to level of need while considering potential victim(s) and timeframe (Logan, 2014).

Several critiques of SPJ measures have been advanced, including their use of clinical judgement to identify and weigh essential risk factors instead of employing numerically generated scores or mathematically derived algorithms (Harris et al., 2015). Moreover, there are no objective criteria with which to classify final risk formulations (i.e. SRRs), and some studies demonstrate the predictive accuracy of SRR judgements to be lower than that of numeric risk scores generated by tallying the items (Belfrage et al., 2012; Desmarais et al., 2012; de Vogel & de Ruiter, 2006; de Vogel, de Ruiter, Hildebrand, Bos, & van de Ven, 2004; de Vogel, de Ruiter, van Beek, & Mead, 2004; Michel et al., 2013; Storey, Kropp, Hart, Belfrage, & Strand, 2014). Inter-rater agreement also tends to be higher for total numeric scores than final risk formulations (Sutherland et al., 2012). These issues concerning the validity and reliability of SPJ SRR judgements may be attributable to a variety of factors including the susceptibility of clinical judgement to decision bias, or alternatively worded, examiner bias (Garb, 1997; Murray, Thomson, Cooke, & Charles, 2011).

The critiques advanced toward the use of SPJ SRRs underscore the need for research that examines how SPJ users arrive at these ratings. Given the lack of explicit criteria to bridge risk item ratings with final formulations in the SPJ tradition, it is critical to better understand this process across raters, both to potentially allay such concerns and/or to offer some explanation for weakened psychometric

properties. Given its established efficacy and frequent usage, the HCR–20:V3 is a natural candidate for which to examine this SPJ issue. There are no studies to our knowledge that examine predictors of SRRs across *multiple raters* for the HCR–20:V3, but there are some for *single raters*.

It has been suggested that the interaction of risk item presence and relevance ratings correlate strongly with SRRs: an association where relevance ratings hold weaker correlations with SRRs than item presence ratings (Howe et al., 2016; Smith et al., 2014). Howe et al. (2016) found that HCR–20:V3 risk factors may have differential correlations with each of the three SRRs (i.e. Case Prioritization, Serious Physical Harm, and Imminent Violence), and that there was low inter-rater agreement about how items and subscales informed subsequent SRRs. In their study, all three SRRs held strong correlations with previous violence and insight, while substance abuse and traumatic experiences presented no association with the same three SRRs. Furthermore, the authors found that Personal Support correlated significantly with the Case Prioritization SRR but not the other two SRRs. More research was recommended for a clearer idea of associations between items and subscales and SRRs, and their implications.

Although research on the item-level predictors of SRR is limited, the early snapshot suggests there may be some issues on the use of SRRs without more explicit rating criteria. For instance, the absence of objective criteria to complete SRRs may result in inconsistencies on what is perceived as a risk factor, despite agreement on overall risk level. Discrepancies between perceived risk factors could result in suboptimal targets for intervention, in addition to difficulties in the communication of risk amongst the treatment team and to decision-makers (Harris et al., 2015; Hilton, Carter, Harris, & Sharpe, 2008). Psychometrically, internal consistency (degree of coherence among test items) is an important property for establishing whether an assessment of a

construct (e.g. violence risk) is reliable. Evaluating predictors of SPJ final risk formulations can help clarify how clinicians use risk factors to understand risk management, determine whether clinicians deviate from standards of use, and develop strategies to optimize completion in the event of extreme deviation.

Current study

The primary aim of this study was to investigate the predictors of SRRs in the HCR–20:V3 across four raters: in other words, how users make the step between risk factor ratings and the final risk formulations. It was a secondary aim to determine the degree to which these predictors coincided with recommended standards of use and empirically based risk factors. These goals will inform how HCR–20:V3 users, and SPJ risk assessment instruments, weigh different sources of information for final risk formulations to guide risk management. Based on the research in support of the psychometric properties of the HCR–20:V3 and latent biases in clinical judgement, it was anticipated that predictors of SRRs should consist mainly of risk item content but may also belie small rater-unique biases.

Method

Design

The study used a retrospective cohort design whereby HCR–20:V3 risk assessments were coded using archived patient information. The sample consisted of randomly selected not criminally responsible on account of mental disorder (NCR) inpatients under warrant in 2016 at a hospital in Western Canada. This project was a part of a province-wide *NCR Project* (Haag, Cheng, & Wirove, 2016): an endeavour that has studied the entire population of NCR individuals throughout the province's history. Extant data on demographic (i.e. sex, race, education and marital status) and criminological variables (i.e. seriousness of index offence and age at NCR verdict) were

drawn from the *NCR Project* database and used for analysis.

Participants

The 32 participants were composed of 26 men and 6 women with a mean age of 37.3 years. Approximately half of the sample completed high school. The racial ancestry of the sample was primarily White (75%), followed by Asian (12.5%), African Canadian (9.4%) and Hispanic (3.1%). For marital status, 59% of the sample were single (i.e. divorced, single, widow) with the remainder (40.6%) married or common-law. Most participants were classified as having a psychotic spectrum disorder (84.3%), followed by mood disorder (28.1%) and/or substance use disorder (56.2%). Most (71.8%) had index violent offences against persons while 28.1% had index non-violent offences.

Measure

Historical Clinical Risk Management-20, Version 3

The HCR-20:V3 (see Douglas et al., 2014, for a detailed review) is an SPJ violence risk assessment tool with 20 empirically supported items associated with violence. Relevance ratings are at each item to allow users to weigh its importance to the case. The ratings for each item are coded on a 3-point scale for both the item's presence (i.e. no, partially present, yes) and the item's relevance (i.e. low, moderate and high). Three SRRs are then rated on a 3-point scale (i.e. low, moderate and high) to assist with case conceptualization and risk management: Case Prioritization/Future Violence, Serious Physical Harm, and Imminent Violence. A two-way mixed interclass correlation procedure revealed high inter-rater reliability across four raters in the current sample for HCR-20:V3 total score (.973), Case Prioritization/Future Violence (.877), Serious Physical Harm (.900), and Imminent Violence (.830; Haag, Hogan, & Cheng, 2017).

Procedure

Four trained raters each coded a HCR-20:V3 using patient file information from 32 participants. The study period ranged from September 2015 to August 2016. One rater was a Registered Psychologist that specialized in forensics, whereas the other three raters were senior undergraduate students under the supervision of the psychologist. All raters received recommended HCR-20:V3 training standards as outlined by the official training manual. Moreover, one of the authors of the HCR-20:V3 was consulted during the project. To ensure consistency, raters compared scores and 'recalibrated' after the first official (non-practice) case and again halfway through the study. Moreover, group consultation was available for uncertainty with risk assessment scoring. Raters operated under the assumption that patients were to be released on an Absolute Discharge at the time of the risk assessment. During the completion of the tool, raters reviewed identical patient information and independently completed a HCR-20:V3 on each participant. The official HCR-20:V3 manual was referenced for each item scored.

Analytic strategy

Two sets of analyses were conducted. First, Pearson's correlation coefficients were computed between SRRs, as well as the individual items and item totals across raters. Median item total correlations were reported with the minimum and maximum item correlates for each SRRs. Second, a series of sequential multiple regressions were conducted across the four raters using nine predictors: the HCR-20:V3 subscales (i.e. Historical, Clinical, Risk Management), demographic variables (i.e. gender, race, marital status, education) and criminological variables (i.e. seriousness of index offence and age at NCR verdict variables). The three SRRs (i.e. Case Prioritization, Serious Physical Harm, Imminent Violence) served as dependent variables. Cohen's f^2 statistic was used to provide local effect sizes for each predictor in the

regressions. Cohen’s (1988) guidelines indicated that $f^2 \geq 0.02$, $f^2 \geq 0.15$, and $f^2 \geq 0.35$ represented small, medium and large effect sizes, respectively.

Ethics

Ethical approval for this project was acquired from the provincial health authority and the University Research Ethics Office.

Results

Bivariate associations between HCR–20:V3 items and SRRs

The results of the correlation analyses are listed in Table 1. The median correlation between item totals and SRRs revealed a

significant positive relationship with large effects across raters. Overall, stability was found across raters for minimum item correlates but greater variation on the maximum end. The two mental disorder items (major mental disorder, symptoms of major mental disorder) were generally among the smallest correlations with SRRs, especially for Case Prioritization for three raters. Minimum correlations for the other two SRRs alternated between the mental disorder items and the traumatic experiences item for the first three raters while the fourth rater had more varied results. For some raters, minimum items had non-significant negative associations with Case Prioritization and Imminent Violence. Although stability was found within raters for maximum item correlates across the three sets

Table 1. HCR–20:V3 item correlations with SRRs.

Median item total and SRR correlation				
	Rater 1 .717**	Rater 2 .762**	Rater 3 .848**	Rater 4 .666**
<i>Case Prioritization</i>				
Minimum	Major mental disorder (.025)	Traumatic experiences (–.071)	Major mental disorder (–.001)	Major mental disorder (–.054)
Maximum	Treatment or supervision response (.842**)	Violent attitudes (.611**)	Other antisocial behaviour (.821**)	Other antisocial behaviour (.672**)
<i>Serious Physical Harm</i>				
Minimum	Traumatic experiences (.095)	Traumatic experiences (.032)	Major mental disorder (.038); symptoms of major mental disorder (.038)	Treatment or supervision response (.045)
Maximum	Treatment or supervision response (.714**)	Violent attitudes (.749**)	Violent ideation or intent (.757**)	Professional services (.567**)
<i>Imminent Violence</i>				
Minimum	Traumatic experiences (.052)	Traumatic experiences (.020)	Symptoms of major mental disorder (–.023)	Stress or coping (.085)
Maximum	Violent ideation or intent (.761**)	Instability (.722**)	Treatment or supervision response (.753**)	Violent ideation or intent (.669**)

Note. SRR = summary risk rating.
**Correlation is significant at the .01 level (two-tailed).

Table 2. Multiple regression of HCR-20:V3 summary risk ratings using subscales and demographics for Rater 1.

Regression model	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	f^2
<i>SRR: Case Prioritization</i>						
Historical ^a	0.118	0.018	0.710	6.726	<0.001	1.008
Risk management ^a	0.074	0.029	0.264	2.503	0.018	0.273
<i>SRR: Serious Physical Harm</i>						
Historical ^b	0.089	0.018	0.556	5.069	0.000	0.669
Seriousness of index offence ^b	-0.334	0.078	-0.460	-4.298	0.000	0.518
Marital status ^b	-0.473	0.168	-0.327	-2.822	0.009	0.346
<i>SRR: Imminent Violence</i>						
Historical ^c	0.071	0.025	0.478	2.799	0.009	0.544
Clinical ^c	0.135	0.057	0.409	2.394	0.023	0.448

Note. SRR = summary risk rating.

^a $R^2 = .848$.

^b $R^2 = .779$.

^c $R^2 = .713$.

of SRRs, between-rater consistency yielded considerable variation. For Case Prioritization and Imminent Violence, two of the four raters had the same item with the maximum association with SRRs; however, they differed in their effect size magnitudes.

Multiple regression predictors of SRRs as a function of rater

Rater 1

Table 2 reports regression findings for HCR-20:V3 subscale, demographic and criminogenic predictors for Rater 1 for the three SRRs. For Case Prioritization, the Historical and Risk Management subscales each predicted global rating. For instance, for every 1-unit increase in Historical score, there was a 0.118-unit increase in Case Prioritization rating and a 0.074-unit increase for each unit advance in Risk Management score. Indeed, 84.8% of the variance in Case Prioritization rating was accounted for by the Historical and Risk Management subscales. In turn, Serious Physical Harm was uniquely predicted by the Historical subscale, seriousness of index offence (i.e. decreasing seriousness predicted lower SRRs), and marital status (i.e. married predicted lower SRRs), in all accounting for

77.9% of the variance. Finally, Imminent Violence was uniquely predicted by each of the Historical and Clinical subscales, accounting for 71.3% of the variance in SRRs.

Rater 2

Table 3 reports regression findings for HCR-20:V3 subscale, demographic and criminogenic predictors for Rater 2 for the three SRRs. For Case Prioritization, the Historical and Risk Management subscales each predicted global rating. For instance, for every 1-unit increase in Historical score, there was a 0.110-unit increase in Case Prioritization rating and a 0.070-unit increase for each unit advance in Risk Management score. Indeed, 76.6% of the variance in Case Prioritization rating was accounted for by the Historical and Risk Management subscales. In turn, Serious Physical Harm was uniquely predicted by the Historical subscale, gender (i.e. being female decreased risk rating) and seriousness of index offence, in all accounting for 62.7% of the variance. Finally, Imminent Violence was uniquely predicted by the Historical subscale, accounting for 43.9% of the variance in SRR.

Table 3. Multiple regression of HCR–20:V3 summary risk ratings using subscales and demographics for Rater 20.

Regression model	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	<i>f</i> ²
<i>SRR: Case Prioritization</i>						
Historical ^a	0.110	0.020	0.676	5.368	<0.001	0.917
Risk management ^a	0.070	0.034	0.259	2.057	0.049	0.268
<i>SRR: Serious Physical Harm</i>						
Historical ^b	0.081	0.021	0.463	3.917	0.001	0.463
Gender ^b	−0.701	0.250	−0.351	−2.803	0.009	0.375
Seriousness of index offence ^b	−0.250	0.100	−0.315	−2.514	0.018	0.332
<i>SRR: Imminent Violence</i>						
Historical ^c	0.125	0.026	0.662	4.842	0.000	0.883

Note. SRR = summary risk rating.

^a*R*² = .7660.

^b*R*² = .6270.

^c*R*² = .4390.

Rater 3

Table 4 reports regression findings for HCR–20:V3 subscale, demographic and criminogenic predictors for Rater 3 for the three SRRs. For Case Prioritization, the Clinical subscale, gender and seriousness of index offence each predicted global rating. For instance, for every 1-unit increase in Clinical score, there was a 0.133-unit increase in Case Prioritization rating, and a 0.392-unit decrease if the participant was female. Indeed, 68.2% of the variance in Case Prioritization rating was accounted for by the Clinical subscale, gender and seriousness of index offence. In turn, Serious Physical Harm was uniquely predicted by the Clinical subscale, age at NCR verdict, seriousness of index offence and race, in all accounting for 83.0% of the variance. Finally, Imminent Violence was uniquely predicted by the Clinical subscale and age at NCR verdict, accounting for 68.6% of the variance in SRR.

Rater 4

Table 5 reports regression findings for HCR–20:V3 subscale, demographic and criminogenic predictors for Rater 4 for the three SRRs. For Case Prioritization, the Historical and Clinical subscale and age at NCR verdict each predicted global rating. For instance, for

every 1-unit increase in Historical score, there was a 0.083-unit increase in Case Prioritization rating and a 0.092-unit increase for each advance in Clinical score. Indeed, 84.4% of the variance in Case Prioritization rating was accounted for by the Historical and Clinical subscale and age at NCR verdict. In turn, Serious Physical Harm was uniquely predicted by the Historical subscale and gender, in all accounting for 53.9% of the variance. Finally, Imminent Violence was uniquely predicted by the Clinical subscale, accounting for 35.2% of the variance in the SRR.

Discussion

The aims of this study were to identify predictors of HCR–20:V3 user SRRs and the degree to which these predictors coincided with recommended standards of use and empirically based risks of future violence. The results would indicate ‘how’ users of the tool use assessment information to arrive at their final SRR judgements. Generally, the correlation and regression analyses suggested that users consistently utilized risk-relevant information for SRR judgements. There was variability, however, it seems, in terms of what information was prioritized in the formulation of such appraisals. That is, consistency was not

Table 4. Multiple regression of HCR-20:V3 summary risk ratings using subscales and demographics for Rater 3.

Regression model	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	f^2
<i>SRR: Case Prioritization</i>						
Clinical ^a	0.133	0.022	0.654	6.074	<0.001	0.865
Gender ^a	−0.392	0.155	−0.293	−2.531	0.017	0.306
Seriousness of index offence ^a	−0.128	0.061	−0.240	−2.088	0.046	0.247
<i>SRR: Serious Physical Harm</i>						
Clinical ^b	0.181	0.023	0.625	7.806	0.000	0.801
Age at NCR verdict ^b	−0.029	0.005	−0.523	−5.892	0.000	0.614
Seriousness of index offence ^b	−0.223	0.063	−0.294	−3.554	0.001	0.308
Race ^b	0.155	0.064	0.216	2.413	0.023	0.221
<i>SRR: Imminent Violence</i>						
Clinical ^c	0.233	0.033	0.728	6.989	0.000	1.062
Age at NCR verdict ^c	−0.023	0.006	−0.378	−3.634	0.001	0.408

Note. SRR = summary risk rating; NCR = not criminally responsible on account of mental disorder.

^a $R^2 = .6820$.

^b $R^2 = .8300$.

^c $R^2 = .6860$.

Table 5. Multiple regression of HCR-20:V3 summary risk ratings using subscales and demographics for Rater 4.

Regression model	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	f^2
<i>SRR: Case Prioritization</i>						
Historical ^a	0.083	0.018	0.515	4.626	<0.001	0.601
Clinical ^a	0.092	0.025	0.386	3.672	0.001	0.418
Age at NCR verdict ^a	−0.013	0.004	−0.230	−2.839	0.008	0.236
<i>SRR: Serious Physical Harm</i>						
Historical ^b	0.109	0.021	0.667	5.273	0.000	0.895
Gender ^b	−0.508	0.244	−0.263	−2.078	0.047	0.273
<i>SRR: Imminent Violence</i>						
Clinical ^c	0.129	0.032	0.593	4.039	0.000	0.737

Note. SRR = summary risk rating; NCR = not criminally responsible on account of mental disorder.

^a $R^2 = .844$.

^b $R^2 = .539$.

^c $R^2 = .352$.

uniform, as there was greater similarity among the items that were least predictive of SRR judgement across raters than those with the largest correlations. At the maximum magnitude for item-level correlations, there was substantial variation in what single items best predicted SRRs. These item-level findings are

consistent with the limited data collected on the HCR-20:V3 by others (Howe et al., 2016; Smith et al., 2014).

The implications for forensic clinical practice may be considerable. Of the individual item predictors in the regressions, no single item consistently and uniformly predicted each

of the three SRR judgements across raters. Moreover, variables that are not explicitly a part of the HCR–20:V3 were predictive of SRRs for some raters (i.e. gender of the patient, age at NCR verdict, marital status and race). The above is viewed as a potential concern given that other variables predicted the final SRR despite the best efforts of the current study to ensure consistency in scoring of the HCR–20:V3, including two calibrations, official training and scoring materials, and group consultation. Taken together this indicates that evaluators using this tool may be considering variables with less relevance to violence risk when completing the SRR.

A further implication for forensic practice is for clinicians who use the HCR–20:V3 (or other SPJ measures) to consider entering the item ratings from their HCR–20:V3 assessments into a data spreadsheet, along with other demographic/clinical variables, and routinely perform regression analyses on their own work to identify the best predictor of SRRs. This practice might help promote scoring fidelity of the HCR–20:V3. From the current study, all raters had strong agreement with final SRRs in terms of inter-rater reliability (Haag et al., 2017); however, close examination elucidated that the raters may consider different pathways to arrive at their SRRs, indicative of their own personal biases. For instance, some raters may have an implicit bias that a serious index offence equates to high potential for future violence (e.g. imminence and severity), while perhaps others may have one of females being lower risk.

A final possible implication from the current findings follows from the above and pertains to risk management. Given that different raters may well arrive at their risk ratings differently, the risk factors that any one clinician may be prioritizing for risk reduction services could, in principle, have some variability. It is unlikely, however, that critical risk factors would be overlooked, given the high inter-rater reliability of the tool and that the entire profile of item presence and relevance ratings would

be available for users to guide intervention planning. For instance, if an individual had active symptoms of psychosis, pro-criminal attitudes and a lack of community supports, both present and relevant to their violence risk, presumably these would still be identified as such by independent raters, although other variables might well factor into the SRR. Although treatment planning is not based on the SRR alone, this would likely inform treatment dosage, release recommendations (e.g. supervision intensity) and the like. Overall, the concerns raised by this study point to the need for more research that examines how SPJ users arrive at final risk formulations.

Strengths and limitations

A strength of this study was the large effect sizes found across most predictors using a relatively small sample size. This finding suggests that the phenomenon of interest, predictors of final risk formulations, was well captured by the variables investigated. A limitation of this study was the insufficient power to determine item-level predictors of SRRs at greater length. Future studies will require large sample sizes to elucidate fine detail as to how clinicians consider individual risk factors, instead of subscales, to produce an overall SRR. The authors recommend large-scale replication studies of the current study involving over 100 participants to allow for further item-level analysis in predicting SRRs. Moreover, an examination of other SPJ tools in the same manner should help reveal the process of clinical judgement between risk factors and SPJ final risk formulations.

Ethical standards

Declaration of conflicts of interest

Jeremy Cheng has declared no conflicts of interest.

Andrew M. Haag has declared no conflicts of interest.

Mark Olver has declared no conflicts of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Acknowledgements

Special acknowledgements are made to Chayse Haldane, Troy Rieck, Alicia MacNeil, Robi Wirove and Arielle Boyes.

Funding

Jeremy Cheng was supported by a Doctoral Fellowship from the Social Sciences and Humanities Research Council of Canada.

References

- Andrews, D. A., & Bonta, J. (1995). *The level of supervision inventory—revised* (Vol. 106, pp. 19–52). Toronto: Multi-Health Systems.
- Belfrage, H., Strand, S., Storey, J. E., Gibas, A. L., Kropp, P. R., & Hart, S. D. (2012). Assessment and management of risk for intimate partner violence by police officers using the Spousal Assault Risk Assessment Guide. *Law and Human Behavior*, 36(1), 60–67 doi:10.1037/h0093948
- Bjorkly, S., Eidhammer, G., & Selmer, L. E. (2014). Concurrent validity and clinical utility of the HCR-20V3 compared with the HCR-20 in forensic mental health nursing: Similar tools but improved method. *Journal of Forensic Nursing*, 10(4), 234–242. doi:10.1097/JFN.0000000000000047
- Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines, and training. *American Psychologist*, 51(9), 945–956. doi:10.1037/0003-066X.51.9.945
- Cohen, J. E. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- de Vogel, V. D., de Ruiter, C. D., Hildebrand, M., Bos, B., & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health*, 3(2), 149–165. doi:10.1080/14999013.2004.10471204
- de Vogel, V., & de Ruiter, C. (2006). Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime & Law*, 12(3), 321–336. doi:10.1080/10683160600569029
- 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.
- de Vogel, V., van den Broek, E., & de Vries Robbé, M. (2014). The use of the HCR-20V3 in Dutch forensic psychiatric practice. *International Journal of Forensic Mental Health*, 13(2), 109–121. doi:10.1080/14999013.2014.906518
- Desmarais, S. L., Sellers, B. G., Viljoen, J. L., Cruise, K. R., Nicholls, T. L., & Dvoskin, J. A. (2012). Pilot implementation and preliminary evaluation of START: AV assessments in secure juvenile correctional facilities. *International Journal of Forensic Mental Health*, 11(3), 150–164. doi:10.1080/14999013.2012.737405
- Douglas, K. S. (2014). Version 3 of the Historical-Clinical-Risk Management-20 (HCR-20V3): Relevance to violence risk assessment and management in forensic conditional release contexts. *Behavioral Sciences & the Law*, 32(5), 557–576. doi:10.1002/bsl.2134
- Douglas, K., & Belfrage, H. (2014). Interrater reliability and concurrent validity of the HCR-20 Version 3. *International Journal of Forensic Mental Health*, 13(2), 130–139. doi:10.1080/14999013.2014.908429
- Douglas, K., Hart, S., Webster, C., Belfrage, H., Guy, L. S., & Wilson, C. M. (2014). Historical-Clinical-Risk Management-20, Version 3 (HCR-20 V3): Development and overview. *International Journal of Forensic Mental Health*, 13, 93–108. doi:10.1080/14999013.2014.906519
- Doyle, M., Power, L. A., Coid, J., Kallis, C., Ullrich, S., & Shaw, J. (2014). Predicting post-discharge community violence in England and Wales using the HCR-20V3. *International Journal of Forensic Mental Health*, 13(2), 140–147. doi:10.1080/14999013.2014.906517

- Farrington, D. P., Jolliffe, D., & Johnstone, L. (2008). *Assessing violence risk: A framework for practice*. Edinburgh, UK: Risk Management Authority Scotland.
- Garb, H. N. (1997). Race bias, social class bias, and gender bias in clinical judgment. *Clinical Psychology: Science and Practice*, 4(2), 99–120. doi:10.1111/j.1468-2850.1997.tb00104.x
- Haag, A., Cheng, J., & Wirove, R. (2016). Describing the not criminally responsible population in Alberta's history: Sociodemographic, mental health, and criminological profiles. *Journal of Community Safety and Well-Being*, 1(3), 68–74.
- Haag, A., Hogan, N., & Cheng, J. (2017). *Evaluating the Interrater Reliability and Concurrent Validity of the HCR-20:V3 with Patients Under the Alberta Review Board*. Unpublished manuscript.
- Harris, G. T., Rice, M. E., Quinsey, V. L., & Cormier, C. A. (2015). Violent offenders: Appraising and managing risk (3rd ed.). Violent offenders: Appraising and managing risk (3rd ed.). (Vol. xviii). Retrieved from <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2014-38766-000>
- Hart, S.D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology*, 3, 121–138. doi:10.1111/j.2044-8333.1998.tb00354.x
- Hilton, N., Carter, A. M., Harris, G. T., & Sharpe, A. J. (2008). Does using nonnumerical terms to describe risk aid violence risk communication? Clinician agreement and decision making. *Journal of Interpersonal Violence*, 23(2), 171–188. doi:10.1177/0886260507309337
- Howe, J., Rosenfeld, B., Foellmi, M., Stern, S., & Rotter, M. (2016). Application of the HCR-20 Version 3 in civil psychiatric patients. *Criminal Justice and Behavior*, 43(3), 398–412. doi:10.1177/0093854815605527
- Kötter, S., von Franqué, F., Bolzmacher, M., Eucker, S., Holzinger, B., & Müller-Isberner, R. (2014). The HCR-20V3 in Germany. *International Journal of Forensic Mental Health*, 13(2), 122–129. doi:10.1080/14999013.2014.911784
- Latimer, J., & Lawrence, A. (2006). *The review board systems in Canada: Overview of results from the Mentally Disordered Accused Data Collection Study*. Ottawa, ON: Department of Justice Canada.
- Logan, C. (2014). The HCR-20 Version 3: A case study in risk formulation. *International Journal of Forensic Mental Health*, 13(2), 172–180. doi:10.1080/14999013.2014.906516
- Michel, S. F., Riaz, M., Webster, C., Hart, S. D., Levander, S., Müller-Isberner, R., ... Hodgins, S. (2013). Using the HCR-20 to predict aggressive behavior among men with schizophrenia living in the community: Accuracy of prediction, general and forensic settings, and dynamic risk factors. *International Journal of Forensic Mental Health*, 12(1), 1–13. doi:10.1080/14999013.2012.760182
- Mossman, D. (2000). Evaluating violence risk “by the book”: A review of HCR-20: Assessing risk for violence, version 2 and the manual for the sexual violence risk—20. *Behavioural Sciences and the Law*, 18, 781–789. doi:10.1002/bsl.418
- Müller-Isberner, R., Webster, C. D., & Gretenkord, L. (2007). Measuring progress in hospital order treatment: Relationship between levels of security and C and R scores of the HCR-20. *International Journal of Forensic Mental Health*, 6(2), 113–121. doi:10.1080/14999013.2007.10471256
- Murray, J., Thomson, M. E., Cooke, D. J., & Charles, K. E. (2011). Influencing expert judgment: Attributions of crime causality. *Legal and criminological psychology*, 16(1), 126–143. doi:10.1348/135532510X490183
- Smith, S. T., Kelley, S. E., Rulseh, A., Sörman, K., & Edens, J. F. (2014). Adapting the HCR-20V3 for pre-trial settings. *International Journal of Forensic Mental Health*, 13(2), 160–171. doi:10.1080/14999013.2014.906520
- Storey, J. E., Kropp, P. R., Hart, S. D., Belfrage, H., & Strand, S. (2014). Assessment and management of risk for intimate partner violence by police officers using the brief spousal assault form for the evaluation of risk. *Criminal Justice and Behavior*, 41(2), 256–271. doi:10.1177/0093854813503960
- 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. *International Journal of Forensic Mental Health*, 11(2), 119–133. doi:10.1080/14999013.2012.690020