



Predictive validity of the HCR-20^{V3} in a sample of Australian forensic psychiatric patients

Delene M. Brookstein^{a,b}, Michael Daffern^{a,b}, James R. P. Ogloff^{a,b}, Rachel E. Campbell^b and Chi Meng Chu^{c,d}

^aCentre for Forensic Behavioural Science, Swinburne University of Technology, Alphington, VIC, Australia; ^bVictorian Institute of Forensic Mental Health (Forensicare), Melbourne, Australia; ^cTranslational Social Research Division, National Council of Social Service, Singapore; ^dMinistry of Social and Family Development, Policy Research Office, Singapore

The Historical Clinical Risk Management-20 Version 3 is the latest iteration in the HCR-20 series, adopting novel changes such as the addition of Relevance ratings and non-requirement to include the Psychopathy Checklist–Revised. This study aimed to examine these changes and compare the predictive validity of the HCR-20^{V3} to the HCR-20^{V2}. The sample comprised of 100 forensic psychiatric patients, retrospectively followed up for a maximum period of approximately 13 years post-discharge from the Thomas Embling Hospital. Recidivism data were sourced from official police records. Results indicated good to excellent inter-rater reliability. The HCR-20^{V3} significantly predicted violent recidivism (area under the curve = .70 to .77), levels of accuracy that were not significantly different from the HCR-20^{V2}. HCR-20^{V3} Relevance ratings failed to add incremental validity above Presence ratings; however, the PCL–R improved upon the HCR-20^{V3}'s validity. The study represented one of the first evaluations of the HCR-20^{V3} in Australia.

Keywords: HCR-20^{V3}; Historical Clinical Risk Management-20 Version 3; PCL–R; predictive validity; Psychopathy Checklist–Revised; reliability; violence risk assessment.

The Historical Clinical Risk Management-20 Version 2 (HCR-20^{V2}; Webster, Douglas, Eaves, & Hart, 1997) is one of the most commonly used violence risk assessment tools amongst mental health professionals (Singh et al., 2014). Revision of the HCR-20^{V2} was underpinned by research developments and contemporary changes in violence risk assessment. As the HCR-20 risk factors are selected based on reviews of the literature as opposed to construction samples, updates to reflect the body of research were required. Notably,

thousands of studies on violence had been published since the release of the HCR-20^{V2} (Douglas et al., 2014).

The revision to HCR-20^{V2}, HCR-20^{V3} (Douglas, Hart, Webster, & Belfrage, 2013) focused on incorporating conceptual developments in the field such as increased emphasis on the decision-making process, formulation and idiosyncratic assessment as a structured professional judgment (SPJ) measure (Douglas et al., 2014). The revision represented a 'reorganization' of risk-relevant

Correspondence: Delene Brookstein, Centre for Forensic Behavioural Science, Swinburne University of Technology, 582 Heidelberg Rd, Alphington, VIC 3078, Australia. E-mail: Delene.Brookstein@forensicare.vic.gov.au

This paper was presented at the International Association of Forensic Mental Health Sixteenth Annual Conference in New York (2016).

information with a focus on continuity of the original concept (Douglas et al., 2013, p. 29). This infers comparable decisions about violence risk factors and probabilities (i.e. provided with the same information, it would be unlikely that the HCR-20^{V2} and HCR-20^{V3} would produce diametrically opposed outcomes; Douglas et al., 2013).

The framework remained stable across versions: 20 risk factors spread diachronically across three scales, capturing past *historical* information (H-scale), present *clinical* considerations (C-scale) and anticipated future contextual *risk management* considerations (R-scale), through use of static and dynamic risk factors. The presence of risk factors is still aptly captured under ‘Presence ratings’ and reflects a judgment on risk factors being conclusively present, partially present or absent. Presence ratings can also be omitted based on lack of reliable information. The gross outcome of the HCR-20^{V3} remains a summary risk rating (SRR) indicating probabilities of low, moderate or high risk of violence. As a SPJ tool, there is no numeric risk estimate or probability, nor any utility in total scores or cut-off scores for interpretation (Smith, Kelley, Rulseh, Sörman, & Edens, 2014). Notably, there are several SRRs for consideration: Future Violence/Case Prioritization, Serious Physical Harm and Imminence of Violence. Future Violence/Case Prioritization refers to the overall risk of violence and infers the level of intervention required to reduce risk. The latter SRRs refer to projections around the severity and imminence of future violence (Douglas et al., 2013).

There were several key changes in the transition from the HCR-20^{V2} to the HCR-20^{V3}, including changes to risk factor labels and content, a ‘broadening’ or ‘narrowing’ of risk factors, and inclusion of item indicators (examples of how risk factors may manifest) and sub-items for complex risk factors to distinguish between different aspects of an item. According to the HCR-20^{V3} manual (Douglas

et al., 2013), six items were broadened, three were narrowed, one was broadened and narrowed in various respects (H8 Traumatic Experiences), and two ‘new’ items were added by drawing on information previously captured under other HCR-20^{V2} items (H2 Other Antisocial Behaviour and H9 Violent Attitudes). Eight items saw no substantive changes.

Broadening is demonstrated under the HCR-20^{V2} H8 (Early Maladjustment), which has been expanded upon in the HCR-20^{V3} to H8 (History of Problems with Traumatic Experiences) and captures trauma across the lifespan. Narrowing is demonstrated in the HCR-20^{V2} C2 (Negative Attitudes), which has become HCR-20^{V3} C2 (Violent Ideation or Intent) to specifically focus on thoughts or plans to perpetrate violence. The most extensive changes were made to the R-scale (Kötter et al., 2014), relative to the number of items broadened/narrowed on a 5-item scale; however the addition of sub-items were seen more extensively on the H-scale and C-scale (i.e. developmental stages, types of relationships/disorders).

Another significant change to the traditional HCR-20 framework was the addition of Relevance ratings. Relevance in this case refers to the ‘extent to which the factor is critical to the evaluator’s formulation of what caused the evaluatee to perpetrate violence and how best to prevent future violence’ (Douglas et al., 2013, p. 50). Relevance ratings therefore allow the assessor to consider the causal importance of risk factors, providing an additional level of analysis that is formulation focused and individualized, emphasizing that risk factors are not equally relevant to all persons who possess them (Monahan et al., 2001). Based on the decision theory of violence, the relevance of a risk factor can be considered in terms of its functional role in motivating, disinhibiting or destabilizing the evaluatee (Douglas et al., 2013).

Unlike the HCR-20^{V2}, the HCR-20^{V3} recommends but does not require administration

of psychopathy measures, Psychopathy Checklist-Revised (PCL-R; Hare, 1991, 2003) or Psychopathy Checklist: Screening Version (PCL:SV; Hart, Cox, & Hare, 1995). The non-requirement for these measures was supported by several studies indicating that the PCL-R did not add incremental predictive validity to the HCR-20 (Campbell, 2007; Douglas & Webster, 1999; Guy, Douglas, & Hendry, 2010). Limited research has evaluated the intersection between the HCR-20^{V3} and these tools. Penney, Marshall and Simpson (2016) found that the PCL-R failed to predict violent outcomes and did not add predictive power when assessed in combination with HCR-20^{V3} dynamic risk factors. Similarly, Hogan and Olver (2016) found that whilst HCR-20^{V3} Presence, Relevance and SRRs significantly predicted inpatient aggression, PCL-R total scores did not. These findings are in contrast to the larger evidence base and several meta-analyses (Edens, Campbell, & Weir, 2007; Leistico, Salekin, DeCoster, & Rogers, 2008; Salekin, Rogers, & Sewell, 1996) indicating a significant association between the PCL-R and prediction of violence.

Psychometric properties of the HCR-20^{V3}

Given that continuity of concept was a key goal of revision, the HCR-20^{V3} sought to retain core aspects of the HCR-20^{V2}, and, therefore, significant associations between the two versions are expected. This has been supported in forensic psychiatric samples, where Douglas and Belfrage (2014) found significant strong correlations between versions for Presence ratings ($r = .90$) and Scale scores ($r = .76$ to $.87$). This finding has been supported by other studies adopting forensic psychiatric samples (Bjorkly, Eidhammer, & Selmer, 2014; de Vogel, de Vries Robbé, van Kalmthout, & Place, 2014) and civil psychiatric samples (Howe, Rosenfeld, Foellmi, Stern, & Rotter, 2016).

Several studies have demonstrated good to excellent inter-reliability outcomes in

forensic psychiatric samples. Douglas and Belfrage (2014) reported good to excellent inter-rater agreement across a range of HCR-20^{V3} ratings including Items, Sub-items, Scale scores, Presence ratings, Relevance ratings and SRRs. Kötter and colleagues (2014) found excellent to almost perfect levels of agreement for HCR-20^{V3} SRRs and good levels of agreement for Scales (intraclass correlation coefficients, ICCs = $.65$ to $.73$). This was encouraging preliminary data given that raters had no previous experience with SPJ risk assessment tools, and assessment was based on case vignettes. Doyle and colleagues (2014) asked experienced staff to independently rate 20 randomly selected patients who were known to them. They found excellent inter-rater reliability for HCR-20^{V3} Total scores (ICC = $.92$) and Scale scores (ICC = $.90$ to $.93$). Interestingly, as with Kötter and colleagues (2014), reliability was greater for the R-scale than the static H-scale. This contradicts other studies that have found the H-scale to be the most reliably rated for both Presence and Relevance ratings (Smith et al., 2014). In their civil psychiatric sample, Howe et al. (2016) found that HCR-20^{V3} SRRs had lower ICCs than Total scores, suggesting ‘strong agreement about the number of risk factors present, [but] less agreement about how the item ratings inform SRRs’ (p. 409). This finding has also been observed in other studies (de Vogel, de Vries Robbé, et al., 2014; Persson, Belfrage, Fredriksson, & Kristiansson, 2017) and is likely to reflect the restricted variance of SRRs in comparison to Total scores.

Pilot research by de Vogel and colleagues (de Vogel, de Vries Robbé, et al., 2014) examined the predictive validity of the HCR-20^{V3} draft version and compared it to that of the HCR-20^{V2}. Researchers conducted a retrospective file review for 86 forensic psychiatric patients discharged from the Van der Hoeven Kliniek in the Netherlands. The predictive validity outcomes for violent recidivism were

reported at one-year, two-year and three-year follow-up periods. Areas under the curve (AUCs) for Total scores at the respective follow-up points were as follows: HCR-20^{V3} (AUCs = .77, .75 and .67) and HCR-20^{V2} (AUCs = .80, .74 and .67). HCR-20^{V3} SRRs (Future Violence/Case Prioritization) were also provided at the follow-up points (AUCs = .72, .67 and .64). All AUCs were significant and did not differ significantly from each other. The study was limited by a small sample and retrospective archival methodology.

Using a prospective method and naturalistic design, Persson and colleagues (2017) evaluated the predictive validity of the HCR-20^{V3} in a sample of 193 forensic psychiatric patients in hospital and correctional settings in Sweden. Assessments were based on patient interviews and file review. Follow-up data were sourced from prison and hospital records and probation services. At the one-year follow-up, the HCR-20^{V3} Total scores (AUC = .79) and SRRs (AUC = .74) both significantly predicted violence. Drawing on a larger sample with prospective methodology, Doyle and colleagues (2014) evaluated the predictive validity of the HCR-20^{V3} in a sample of 387 forensic psychiatric patients in England and Wales. Patients were sourced from 32 medium-security forensic psychiatry units, from where they had been discharged to community and non-forensic placements. The HCR-20^{V3} was completed at both 6 months and 12 months post-discharge, based on clinical records and interviews with social supervisors and/or care coordinators. Incidents of violence were recorded through clinical and police records and interviews with staff. Results showed that the HCR-20^{V3} Total scores significantly predicted violence at 6-month and 12-month follow-ups (AUCs = .73 and .70), as did the H-scale, C-scale and R-scales at 6-month follow-up (AUCs = .63, .74 and .67) and 12-month follow-up (AUCs = .63, .70 and .63). Violent participants scored significantly higher on the HCR-20^{V3} Total

score at both follow-up points, with those scoring above the median (*Mdn* = 23) being 2–5 times more likely to be violent during follow-up than those scoring below. The study was advantaged by a large sample and interview data in addition to file records.

In the first evaluation of the HCR-20^{V3} in a Scottish sample, Smith and colleagues (Smith, O'Rourke, & Macpherson, 2020) found that the HCR-20^{V3} Total scores significantly predicted inpatient violence with forensic psychiatric patients (AUC = .69). The C-scale and R-scale both significantly predicted violence (AUCs ranging from .64 to .72), whereas the H-scale did not (AUC = .51, $p > .05$). The C-scale emerged as the strongest predictor. The authors note that due to the characteristics of forensic psychiatric populations, participants are likely to endorse most historical items, translating to reduced variance for this scale. Likewise, dynamic variables by their changing nature have more scope for variation and hence may emerge as superior predictors.

There is very limited research to inform the predictive value afforded by Relevance ratings. Hopton and colleagues (Hopton, Cree, Thompson, Jones, & Jones, 2018) found that the quality of risk formulations were significantly higher for the HCR-20^{V3} than the HCR-20^{V2}, which may speak to the value of Relevance ratings in further tailoring risk formulations. Strub, Douglas and Nicholls (2014) found that Relevance ratings significantly predicted violent recidivism in the short term (4–6 weeks) and long term (6–8 months); however, they failed to add incremental validity to Presence ratings on Scales and Total scores.

Hogan and Olver have demonstrated the predictive validity of the HCR-20^{V3} for forensic psychiatric samples in both inpatient (Hogan & Olver, 2016) and community settings (Hogan & Olver, 2019). In their community study, they found that post-treatment HCR-20^{V3} Presence and Relevance ratings and SRRs (Future Violence/Case Prioritization) all significantly predicted violent recidivism. Relevance ratings (AUC = .83) outperformed

Presence ratings (AUC = .81) and SRRs (AUC = .73). The study also evaluated incremental predictive validity afforded by change scores in dynamic risk factors (i.e. C-scale and R-scale) pre- and post-treatment. While controlling for time at risk, they found that Relevance rating change scores added significant incremental validity to prediction of violent recidivism, over and above pre-treatment Relevance ratings. This finding was not replicated for Presence ratings, consistent with research by Mastromanno and colleagues (2018) who also found that change scores for dynamic risk variables did not significantly predict violent recidivism.

In summary, preliminary evaluations of the HCR-20^{V3} are promising, enabling implementation of the tool in practice as the evidence base develops. Research within forensic psychiatric, civil psychiatric and/or correctional populations indicate that the HCR-20^{V2} and HCR-20^{V3} are strongly correlated, and that the HCR-20^{V3} demonstrates good to excellent inter-reliability outcomes. Predictive validity indices observed have been towards the upper levels of accuracy reported in the literature. Little is known about how the selective exclusion of the PCL-R will impact the validity of HCR-20^{V3} violence risk assessments; however, preliminary research has indicated that personality disorder is being scored differently on the HCR-20^{V2} and HCR-20^{V3}, possibly due to the less stringent criteria in assessing psychopathy on the HCR-20^{V3} (Smith et al., 2014). There has been limited exploration of the Relevance ratings as a novel element of the tool. Studies suggest that Relevance ratings can be reliably rated and serve to improve upon risk formulations (Douglas & Belfrage, 2014; Hopton et al., 2018); however, little is known about their predictive capabilities and how they interact with Presence ratings.

The current study aimed to evaluate the predictive validity of the HCR-20^{V3} and compare it to that of the HCR-20^{V2}. Based on

preliminary research and the goals of revision, it was hypothesized that the HCR-20^{V2} and HCR-20^{V3} would demonstrate similar predictive capabilities that would not differ significantly. We also sought to explore the incremental validity afforded by the Relevance ratings and PCL-R. It was hypothesized that Relevance ratings would demonstrate incremental predictive validity over the Presence ratings, and that the PCL-R would not add to the predictive validity of the HCR-20^{V3} Total scores or SRRs.

Method

Setting

The Victorian Institute of Forensic Mental Health (VIFMH, Forensicare) is the state-wide statutory authority for the provision of forensic mental health services in Victoria, Australia (Victorian Institute of Forensic Mental Health, 2015). Governed under the Mental Health Act, 2014 (VIC) (Mental Health Act, 2014), Forensicare provides clinical services across several sites including the Thomas Embling Hospital (TEH), the state's only forensic hospital. Patients include prisoners transferred for involuntary mental health treatment (i.e. security patients), individuals found not guilty by reason of mental impairment (NGRMI; i.e. forensic patients) or involuntary civil psychiatric patients. Notably, the legal designation of patients can change during the course of their admission.

Sampling procedures

Participants were identified through comparative datasets from previous research conducted at the Centre for Forensic Behavioural Science (CFBS), namely Campbell (2007) and Chu (2010). As part of these research projects, participants had received HCR-20^{V2} and PCL-R assessments based on file review during TEH admissions between April 2000 and December 2010. HCR-20^{V2} Scale scores and Total

scores, and PCL-R Total scores and Facet scores were provided. HCR-20^{V2} SRRs were not provided.

The combination of the Campbell (2007) and Chu (2010) datasets resulted in a pool of $N = 186$ participants. The sample was refined through removal of matching cases ($n = 13$), participants identified in the National Coronial Information System (NCIS; $n = 11$), participants who had been deported or housed in secure extended care mental health facilities ($n = 19$), and participants for whom inpatient files were unable to be retrieved from archive ($n = 19$). Of the remaining pool, a random sample of 100 participants were selected for participation in the current study.

Research design

A retrospective file review of 100 adult forensic psychiatric patients hospitalized at TEH was conducted. Following psychiatric treatment, participants were discharged directly into the community or transferred to prison, with delayed community entry. For participants discharged directly into the community, the extent of care was limited to supervision in the community under an Area Mental Health Service, placement at an open-door residential mental health unit or placement in supported accommodation (e.g. residences with additional support for persons with mental illness or disability).

Violence was defined as ‘actual, attempted, or threatened infliction of bodily harm on another person’ (Douglas et al., 2013, p. 36), as per the HCR-20^{V3} operational definition of violence. Victoria Police provided recidivism data in the form of criminal charges extracted from the Law Enforcement Assistance Program (LEAP) database, for the period April 2000–January 2013, enabling a maximum follow-up period of 12 years and 10 months. Violent offences included the following categories with noted examples: homicide offences (murder, manslaughter, culpable driving), sex offences (rape, sexual

penetration, exposure, indecent acts), assault (recklessly/intentionally cause injury, assault with weapon), fire-setting and arson, kidnapping (abduction, false imprisonment, hold against will, unlawfully detain), stalking (stalk/harass persons), threat offences (e.g. use threatening words, threat to kill, extortion with threats) and theft offences with a violent component (e.g. robbery, armed robbery or aggravated burglary with person present).

Due to the serious psychological harm component of the HCR-20^{V3} definition of violence (Douglas et al., 2013, p. 36), stalking and kidnapping offences were coded as violent. Arson and fire-setting were also captured under the definition of violence as they may invoke both physical and psychological harm. Similarly, theft offences with a person present were coded as violent due to risk of psychological harm stemming from fear of physical injury. Burglary did not meet this criterion (i.e. does not necessarily involve person present). Robbery offences were classified as violent due to their definitions in the Crimes Act (1958) in seeking to or actually subjecting another person to use of force. Possession of a regulated/unregistered weapon or unsafe carrying of a weapon was not in itself considered a violent offence.

During the follow-up period, movements in and out of psychiatric hospitals and prisons were provided by the Department of Justice Prisoner Information Management System (PIMS) and Victoria’s Client Management Interface (CMI) database. This enabled the calculation of a ‘time at risk’ variable (total time spent in the community excluding periods of hospitalization or incarceration). Ethics approval for the project was received from Swinburne University Human Research Ethics Committee, Department of Justice Human Research Ethics Committee, Victoria Police Research Coordinating Committee and the National Coronial Information System. A confidential inquiry approach was used as a means of attaining participants thought to be at greater risk of non-compliance, anti-sociality and violence (Doyle et al., 2014).

Measures & scoring

HCR-20^{V3}

The HCR-20^{V3} was retrospectively coded based on file review at the point of discharge by a single rater. The rater was a doctoral candidate in clinical and forensic psychology with previous experience in completing HCR-20^{V3} assessments as part of course practicum requirements. Formal training in the administration of the HCR-20^{V3} was completed. Typical sources of file information included comprehensive bio-psycho-social assessments, intensive case reviews, clinical notes and discharge summaries completed by members of multidisciplinary teams. In total, 875 files were reviewed, with an average of 8.75 files per participant (min = 1, max = 52).

All relevant historical information was considered for the H-scale. The HCR-20^{V3} manual notes that evaluators should determine a specific timeframe for coding the C-scale. Guidance is provided around an optimal timeframe that is greater than one month but less than six months (Douglas et al., 2013). The decision regarding timeframes needed to consider feasibility of review and instating a period of time that is comprehensive enough to enable the capturing of clinical information prior to discharge. Given that the average length of stay of acute patients at TEH was 74 days, it was considered that information gathered 2 months prior to discharge would provide sufficient clinical information and be available for most participants (Victorian Institute of Forensic Mental Health, 2012). The R-scale was scored at point of discharge for the foreseeable six months as if the participant was ‘out’ in the community. Scoring at the point of discharge allowed for capturing of contextual considerations (e.g. discharge location, professional and personal supports).

The HCR-20^{V3} ratings were scored blind to outcome. Presence ratings, Relevance ratings and the SRR (Future Violence/Case Prioritization) were coded. As the HCR-20^{V3} Presence and Relevance ratings are scored on a nominal system, an ordinal scale was created

for this study by transposing ratings to numerical scores where, for Presence ratings, 0 = no, 1 = partial/possible, and 2 = yes/definite. For Relevance ratings, 0 = low, 1 = moderate, and 2 = high. Therefore, the Presence and Relevance Ratings Total scores ranged from 0 to 40. For the SRR, 1 = low, 2 = moderate, and 3 = high.

A post-doctoral clinical and forensic psychologist with formal HCR-20^{V3} training assisted with reliability scoring for the HCR-20^{V3}. Ten percent of the sample were randomly selected for reliability scoring. Inter-rater reliability was assessed using the ICC. Table 1 displays the ICCs for the HCR-20^{V3} (Presence Total scores and Scale scores, Relevance Total scores and SRRs). Reliability results were interpreted based on Fleiss (1981) categorizations, where ICCs < .40 were considered ‘poor’, .40 to .59 were ‘moderate’, .60 to .74 were ‘good’, and ≥ .75 were ‘excellent’.

HCR-20^{V2}

It was essential that the HCR-20^{V2} and HCR-20^{V3} ratings occurred at the same time points. Whilst HCR-20^{V2} scores in the Campbell (2007) dataset were coded at discharge (consistent with the current study), the Chu (2010)

Table 1. Inter-rater reliability of the HCR-20^{V3}.

Measure	Inter-rater reliability		
	ICC ₁	ICC ₂	95% CI
HCR-20 ^{V3}	.82	.90	[.44, .95]
Total score			
H-scale	.83	.91	[.46, .95]
C-scale	.68	.81	[-.10, .96]
R-scale	.52	.68	[-.57, .92]
HCR-20 ^{V3} SRR	.68	.81	[.12, .91]
Relevance	.56	.72	[-.11, .87]
ratings total			
score			

Note: HCR-20^{V3} = Historical Clinical Risk Management-20 Version 3; ICC₁ = single measure intraclass correlation coefficient; ICC₂ = average-measure intraclass correlation coefficient; CI = confidence interval; SRR = summary risk rating.

dataset included C-scale and R-scale scores that were coded on admission. Therefore, a research assistant who was a doctoral candidate at the CFBS reviewed the files in order to code the HCR-20^{V2} C-scale and R-scale scores on discharge, consistently with the ranges adopted in the current study.

Participants were omitted from the HCR-20 sample when a threshold of missing data was exceeded (i.e. no more than two items from the H-scale, and one item from the C-scale and R-scale, respectively). One participant from the HCR-20^{V2} dataset was omitted on this basis; however, no participants from the HCR-20^{V3} dataset were omitted. As scores on the original HCR-20^{V2} datasets were prorated in the event of missing scores, the HCR-20^{V3} scores were also prorated according to the same formula to ensure consistency in the datasets. When scores were missing but not to the extent that the threshold of missing data was met, prorated scores were computed.

PCL-R

PCL-R scores from both the Campbell (2007) and Chu (2010) datasets were coded retrospectively based on historical file information, and therefore did not require re-coding. PCL-R items were coded on a 3-point scale where Total scores ranged from 0 to 40. Based on exclusionary criteria, no more than 5 of the 20 items could be omitted, with no more than two items per factor. Due to missing scores, $n = 95$ for the PCL-R dataset.

Participants

The sample comprised 100 forensic psychiatric patients ($n = 73$ males; $n = 27$ females). Mean age at discharge was 33.51 years ($SD = 10.20$; range = 18.51–63.18). Most participants were Australian born ($n = 78$). The remainder of the sample were born in Asia (7%), Europe (8%), New Zealand (5%) or the United Kingdom (2%). The majority (74%) were of English-speaking background, 21% were Culturally and Linguistically Diverse

(CALD), and 5% were of Aboriginal and Torres Strait Islander backgrounds.

Participants were discharged directly into the community (63%) or transferred to prison (37%; i.e. approximately one third of the sample transferred directly from hospital to prison and therefore had delayed community entry). Discharges occurred from acute (83%), sub-acute (11%) and rehabilitation (6%) units. The mean length of inpatient stay was 257.99 days ($Mdn = 59$, $SD = 615.82$; min = 5, max = 3037). On discharge, the legal status of participants was as follows: security patients (72%), involuntary patients (20%) and forensic patients (8%).

Most participants (88%) had some form of employment history. Just over half of the sample had a history of self-harm (51%) and/or a history of suicide risk (60%). Three quarters (77%) had previous civil psychiatric admissions, and more than half (55%) had prior forensic psychiatric hospital admissions. Ninety-five percent of the sample were diagnosed with a mental illness upon discharge from TEH (excluding personality disorder). Primary diagnoses included: schizophrenia spectrum and other psychotic disorders (72%); depressive disorders (8%); bipolar and related disorders (6%); substance related disorders (17%); trauma and stress-related disorders (5%); obsessive compulsive disorder (1%); and eating disorder (1%).

Thirty percent of the total sample was diagnosed with a personality disorder upon discharge. Primary diagnoses included: antisocial personality disorder (13%); borderline personality disorder (11%); paranoid personality disorder (4%); narcissistic personality disorder (2%); and personality disorder not otherwise specified (NOS) (4%). Obsessive-compulsive, schizotypal, dependent and avoidant personality disorders were each represented in 1% of the total sample.

Two thirds (66%) of the sample had previous recorded convictions for violent offences (excluding the index offence). Index offence(s) were defined as offences for which the

participant had been charged or convicted that directly contributed to the participant’s admission to TEH or the current prison term. Index offence(s) included at least one violent offence for 67% of the sample.

Data analysis

As most variables in the current study violated assumptions of normality, non-parametric statistical methods such as Kendall’s tau were employed for correlational analyses. Kendall’s tau is better suited to ordinal scales (Hanley & McNeil, 1983; as were created in the current study for HCR-20^{V3} SRRs) and is considered a superior estimate of actual correlations in the population, leading to more accurate generalizations (Field, 2009; Howell, 1997). For consistency in comparing correlations, Kendall’s tau was used for both SRR and Total score analyses. Pearson’s correlation were reported where available, for ease of comparison with the extant literature.

The receiver operating characteristic (ROC) statistic and resulting AUC were used to evaluate predictive validity. While there is no formal categorization for the interpretation of AUCs and some variation within the literature, the following AUC thresholds were adopted in the current study based on an overview of the literature (Dolan & Doyle, 2000; Douglas & Webster, 1999; Rice & Harris, 2005): less than .65 = small effect; .65 to .70 = moderate effect; and >.70 = large effects.

Other indices of discrimination, such as positive predictive power (PPP) and negative predictive power (NPP) were provided to describe the trade-off between sensitivity and

specificity, and false-positive and false-negative errors (Martinez-Camblor, Carleos, & Corral, 2013; Mossman, 2013). To calculate the positive predictive value (PPV) and negative predictive value (NPV), tools must be treated as though they produced a dichotomous outcome; however, the HCR-20^{V3} produces discrete categories of low, moderate or high risk. Therefore, dichotomous outcomes were created by grouping SRRs as follows: low versus (moderate and high), and (low and moderate) versus high. This binning strategy is consistent with previous research comparing the predictive accuracy of various risk assessment tools where dichotomous outcomes are not readily available (Singh, Grann, & Fazel, 2011). The cut-off score represents the minimum value constituting a test-positive result. The indices of discrimination are displayed in Table 2. There is no utility in reporting indices for the low-risk category, because all cases would be classified as test positive. Furthermore, PPV and NPV are not reported for the HCR-20^{V3} Total scores as this defeats the purpose of the SPJ tools and encourages improper use (Guy, 2008), as the HCR-20^{V3} is not designed to be used based on cut-off scores in this manner.

Hanley and McNeil’s (1983) nonparametric method for comparing AUCs that have been derived from the same participants was employed for the HCR-20^{V3} and HCR-20^{V2} comparison, to account for the paired nature of the data (DeLong, DeLong, & Clarke-Pearson, 1988). Although calculations were computed by hand, MedCalc Software (2016; a statistical software package that allows for pairwise

Table 2. Indices of discrimination based on HCR-20^{V3} SRR categories.

Cut-off dichotomy	Sensitivity	Specificity	PPV	NPV
Low vs. (moderate & high)	1.00	.32	.60	1.00
(Low & moderate) vs. high	.68	.76	.74	.70

Note: N = 100. HCR-20^{V3} = Historical Clinical Risk Management-20 Version 3; PPV = positive predictive value; NPV = negative predictive value.

comparisons of dependent ROC curves) was used to verify the results, as adopted in other studies (Persson et al., 2017).

Survival analysis (Kaplan–Meier method) was used to examine the hypothesis that there would be statistically significant differences in survival time for violent recidivism between HCR-20^{V3} risk categories (SRRs). Survival analysis deals with the amount of time until a specific event occurs and is therefore known as a ‘time to event’ analysis. In the current study, this event is time until being charged with the first violent offence. Participants who were not charged with violent offences during the follow-up period are considered to have ‘survived’, whilst those who were deemed as having ‘failed’.

The research was conducted in accordance with the Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE; Singh, Yang, & Mulvey, 2015). All data analyses were conducted using IBM’s Statistical Package for the Social Sciences (SPSS; Version 22).

Results

The average time at risk for violent offending was 1935.19 days (approximately 5 years and 3 months; *Mdn* = 1397.50, *SD* = 1622.03, *min* = 1, *max* = 4502). The base rate for violent offending was 50%. This unusually high base rate is possibly a reflection of a high-risk sample and/or the longitudinal follow-up period. HCR-20^{V3} Total scores for Presence ratings ranged from 8 to 40, with a high average score (*N* = 100, *M* = 28.47, *Mdn* = 30, *SD* = 6.09). The sample SRRs were as follows: low (*n* = 16), moderate (*n* = 38) and high (*n* = 46). The HCR-20^{V2} Total scores ranged from 2.0 to 37.89 (*n* = 99; *Mdn* = 24, *SD* = 7.12). The average PCL–R Total score was 14.91 (*n* = 95, *Mdn* = 15, *SD* = 6.59).

Kaplan–Meier survival analysis was used to observe differences in violent recidivism patterns over time, associated with the HCR-20^{V3} SRRs. The median survival time for recidivists was 1473 days, 95% confidence

interval (CI) [114.29, 2831.72]. The average survival times in days across SRR groups were as follows: low (*M* = 2597); moderate (*M* = 2199.63) and high (*M* = 1486.54). Figure 1 displays Kaplan–Meier survival curves for HCR-20^{V3} SRRs and time to first violent offence. No participants rated as low risk of violence were charged with a violent offence over the entire follow-up period. Groups were compared using the log rank test. Groups differed significantly in survival distributions, $\chi^2(2) = 21.694$, $p \leq .001$. Low differed significantly from moderate ($\chi^2 = 6.942$, $p = .008$) and high ($\chi^2 = 15.881$, $p = .000$); and moderate differed significantly from high ($\chi^2 = 8.480$, $p = .004$).

Predictive validity of the HCR-20^{V3}

Kendall’s tau correlations indicated significant associations between violent recidivism and the HCR-20^{V3} Total scores ($\tau = .29$, $p < .01$) and SRRs ($\tau = .49$, $p < .01$). Table 3 summarizes the predictive performance of the HCR-20^{V3}. The HCR-20^{V3} SRRs produced the largest AUC of .77 (*SE* = .05, $p = .000$), followed by the Relevance ratings (AUC = .71, *SE* = .05, $p = .000$) and Total scores (AUC = .70, *SE* = .05, $p = .001$). The odds of being charged with a violent offence during the follow-up period was 2.89 times higher for participants scoring above the median Total score (*Mdn* = 30).

Comparing HCR-20^{V3} to HCR-20^{V2}

Internal consistency estimates indicated that the overall reliability of the HCR-20^{V3} was good (Cronbach’s $\alpha = .82$). The reliability of the individual scales was as follows: H-scale Cronbach’s $\alpha = .61$; C-scale Cronbach’s $\alpha = .54$; and R-scale Cronbach’s $\alpha = .74$. The overall reliability of the HCR-20^{V2} was also good (Cronbach’s $\alpha = .79$); however, reliability within individual scales varied greatly: H-scale Cronbach’s $\alpha = .14$; C-scale Cronbach’s $\alpha = .99$; and R-scale Cronbach’s $\alpha = .84$. The HCR-20^{V2} and HCR-20^{V3} Total scores were significantly correlated ($\tau = .53$, $p < .01$) and

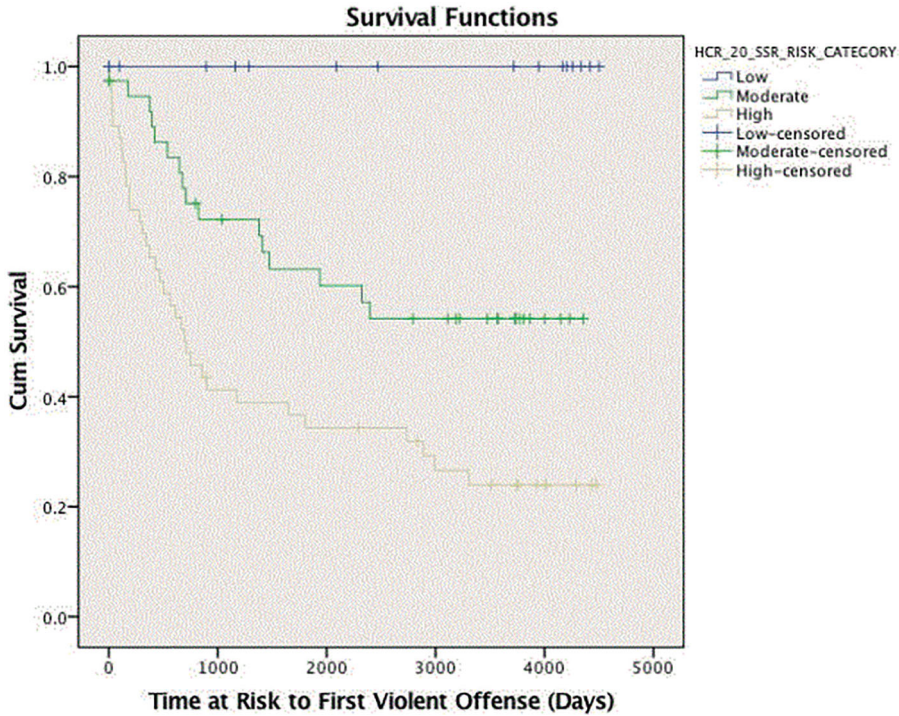


Figure 1. Kaplan–Meier survival curves for HCR-20^{V3} SRRs and time to first violent offense.

Table 3. Predictive performance of the HCR-20^{V3} with violent recidivism as outcome.

Measure	AUC	Effect size	95% CI
HCR-20 ^{V3} Total score	.70	Medium	[.59, .80]
HCR-20 ^{V3} SRR	.77	Large	[.68, .86]
HCR-20 ^{V3} Relevance ratings	.71	Large	[.60, .81]

Note: $N=100$. HCR-20^{V3} = Historical Clinical Risk Management-20 Version 3; SRR = summary risk rating; AUC = area under the curve; CI = confidence interval.

Pearson’s correlation ($r = .74, p < .01$). Scale scores across the two versions also correlated significantly with moderate to good strength: H-scale ($r = .66, p < .01$), C-scale ($r = .63, p < .01$) and R-scale ($r = .50, p < .01$). Due to one case of missing data on the HCR-20^{V2} dataset, $n = 99$.

Separate ROC curves were run for HCR-20^{V2} and HCR-20^{V3} Total scores. The HCR-20^{V2} produced an AUC of .77

($SE = .046, p < .001$). The HCR-20^{V3} produced an AUC of .69 ($SE = .053, p < .01$). The indices of discrimination are displayed in Table 4. Based on these results, the critical z-ratio was calculated to test the null hypothesis that the difference between areas is random. The critical z ratio was below a cut-off of $z \geq 1.96$ ($z = 1.60$), indicating that the true ROC areas were not significantly different.

Table 4. Indices of discrimination for the HCR-20^{V2} and HCR-20^{V3} with violent recidivism as outcome.

Measure	AUC	SE	p	95% CI
HCR-20 ^{V2} total score	.77	.046	.000	[.682, .864]
HCR-20 ^{V3} total score	.69	.053	.001	[.588, .794]

Note: $N=99$. HCR-20^{V3} = Historical Clinical Risk Management-20 Version 3; AUC = area under the curve; CI = confidence interval.

Incremental validity

Relevance ratings

A sequential logistic regression was conducted to assess the incremental validity associated with the Presence and Relevance features of the HCR-20^{V3}. Assumptions of linearity and multicollinearity were met. Preliminary analyses demonstrated that the HCR-20^{V3} Presence and Relevance ratings were moderately correlated ($\tau = .66, p < .01$). Presence ratings were entered into Block 1, producing a significant model with good fit, $\chi^2(1) = 12.51, p < .001$. The model explained 15.7% (Nagelkerke R^2) of variance in violent recidivism and correctly classified 59% of cases (sensitivity was 70%, specificity was 48%). The Wald test indicated that the HCR-20^{V3} Presence ratings were significant predictors of violence (Wald = 10.07; $p < .01$). Relevance ratings were added in Block 2, also resulting in a significant model with good fit, $\chi^2(2) = 14.372, p < .01$. The model explained 17.8% (Nagelkerke R^2) of variance in violent recidivism and correctly classified 67% of cases (sensitivity was 72%, specificity was 62%). The Wald test (Wald $Z = 1.819; p > .05$) indicated that Relevance ratings made non-significant improvements in prediction and did not demonstrate significant incremental validity over the Presence ratings. When considered simultaneously in Block 2, Presence and Relevance ratings both emerged as non-significant predictors.

PCL-R

A sequential logistic regression was conducted to assess the incremental validity associated

with HCR-20^{V3} and PCL-R scores in predicting violent recidivism. Assumptions of linearity and multicollinearity were met. The reliability of the four PCL-R Facet scores was acceptable, whereby Cronbach's $\alpha = .73$. The HCR-20^{V3} Total scores were entered at Block 1, producing a significant model, $\chi^2(1) = 11.436, p < .01$. The percentage of variance explained by the HCR-20^{V3} Total scores was 15.1% (Nagelkerke R^2). When the PCL-R Total scores were added at Block 2, the model remained significant, $\chi^2(1) = 22.403, p < .001$. The contribution of the PCL-R Total scores was significant, $\chi^2(1) = 10.967, p \leq .01$, emerging as the only significant predictor (Wald $Z = 8.98, p < .01$). The percentage of variance in outcome explained by the HCR-20^{V3} Total scores and PCL-R in combination was 28% (Nagelkerke R^2).

The analysis was re-run using the HCR-20^{V3} SRRs, coded as nominal variables. The HCR-20^{V3} SRR was entered at Block 1, producing a significant model with good fit, $\chi^2(2) = 32.107, p < .001$. The percentage of variance explained by the HCR-20^{V3} SRR was 38.3% (Nagelkerke R^2), with 71.6% of cases correctly classified. PCL-R Total scores were added at Block 2. The model remained significant with good fit, $\chi^2(3) = 38.884, p < .001$. The contribution of the PCL-R Total scores was significant, $\chi^2(1) = 6.714, p \leq .05$, again emerging as the only significant predictor (Wald $Z = 5.849, p < .05$).

Discussion

The current study aimed to evaluate the predictive validity of the HCR-20^{V3} and compare it to that of the HCR-20^{V2}. The study also endeavoured to explore the incremental

validity afforded to the HCR-20^{V3} Presence ratings by Relevance ratings, and the incremental validity afforded to the HCR-20^{V3} Total scores and SRRs by the PCL-R.

Association and inter-rater reliability

Results indicated that the HCR-20^{V2} and HCR-20^{V3} were significantly correlated, a finding consistent with previous research in forensic psychiatric samples (de Vogel, de Vries Robbé, et al., 2014; Douglas & Belfrage, 2014; Douglas et al., 2014). This is expected given the ‘continuity of concept’ goal of revision. At the Scale level, the H-scale produced the strongest association whilst the R-scale produced the weakest association, which may be reflective of this scale having undergone the most extensive changes.

Inter-rater reliability for the HCR-20^{V3} Scales ranged from moderate to excellent, with the greatest agreement observed for the H-scale and the least agreement for the R-scale. This is consistent with other studies (Smith et al., 2014), but also contradicts studies that found superior agreement for the R-scale (Doyle et al., 2014). This finding may be explained by the static nature of the H-scale, which contrasts with the prospective nature of the R-scale, involving forecasting of future living and contextual factors that may be more subjective. Levels of agreement were excellent for HCR-20^{V3} Total scores and good for SRRs. This is consistent with previous research demonstrating greater ICCs for Total scores than SRRs (de Vogel, de Vries Robbé, et al., 2014; Douglas et al., 2014; Howe et al., 2016; Persson et al., 2017), but contradicts Penney and colleagues (2016) who found superior ICCs for SRRs. The current results indicate strong agreement about the presence of risk factors, but differences in how item-level risk information culminates in the overall SRR. Indeed, agreement for HCR-20^{V3} Relevance ratings were moderate, which suggests differences in how raters formulated the causal and functional relevance of risk factors, despite the addition of item indicators to

improve consistency. Strub and colleagues (2014) have posed the question of whether clinicians are using and defining Relevance ratings in a consistent manner, or whether the relevance of certain items is weighted differently (see Dickens & O’Shea, 2017).

Predictive validity

The HCR-20^{V3} Total scores and SRRs significantly predicted violent recidivism, producing AUCs in similar ranges to what has been reported in other studies (de Vogel, de Vries Robbé, et al., 2014; Doyle et al., 2014; Persson et al., 2017). The largest AUC was observed for the SRRs, followed by Relevance ratings and finally Presence ratings. These results are similar to previous research identifying SRRs as a robust predictor and indicating the superiority of SRRs over Total scores (Guy, 2008; Strub et al., 2014); however, they contradict other HCR-20^{V3} studies (de Vogel, de Vries Robbé, et al., 2014; Persson et al., 2017). The emergence of SRRs as the strongest predictor in the current study supports the revision goal of embodiment of the SPJ approach in moving away from an actuarial score-focused assessment towards a formulation-driven one.

Although the HCR-20^{V2} and HCR-20^{V3} produced differing AUCs, these were not found to differ significantly from one another, consistent with findings by de Vogel and colleagues (de Vogel, de Vries Robbé, et al., 2014). Discrepancy in observed AUCs may suggest non-significant differences in the discriminant ability of these tools. These results need to be interpreted with caution as the small sample size certainly would have impacted the ability to detect significant differences between ROC curves. Hanley and McNeil (1982) have recommended much larger sample sizes for ROC curves within the ranges of those described in this study. Whilst these guidelines are not binding, they do highlight the lack of statistical power in the current sample.

The indices of discrimination (see Table 2) indicated that the HCR-20^{V3} performed well in identifying the presence and absence of violence. At the high-risk cut-off, the probability of a participant with a positive test being violent was 74%, and the probability of a participant with a negative test not being violent was 70%. At the moderate-risk cut-off, the probabilities were 60% and 100%, respectively. Erring on the side of caution, a reasonable balance was achieved between identifying the presence or absence of violence. These probabilities demonstrate the need for a commensurate level of intervention being provided at various levels of risk (as per the risk–need–responsivity, RNR, model; Bonta & Andrews, 2007).

The pattern of the survival curves across risk groups for the HCR-20^{V3} SRRs differed significantly and produced an interesting survival plot demonstrating exponential curves for the moderate- and high-risk groups. The divergence across risk groups as displayed in Figure 1 supports the use of SRRs, for not only the presence of violence but also timing, as those deemed high risk had the quickest ‘drop-off’, whereas a more gradual slope was observed for the moderate-risk group during the longitudinal follow-up.

While these results are encouraging, it is worth noting that the typical use of the HCR-20 in the context of forensic mental health treatment at TEH is not for diagnostic screening purposes (where high levels of sensitivity are desirable), but rather for prognostic purposes to inform clinical decision making around risk mitigation (Singh et al., 2011). Notably, the focus for security patients transferring back to custodial settings would be on treatment of mental illness (i.e. clinical risk factors); whereas forensic patients gradually progress through TEH from acute to rehabilitation units, with significant input to treatment of mental illness and discharge planning (i.e. clinical and risk management factors). The proportion of security or forensic patients in forensic psychiatric samples may explain

differences in significance of dynamic change scores in studies using pre- and post-treatment designs (see Hogan & Olver, 2019; Mastromanno et al., 2018).

Incremental validity

The hypothesis that HCR-20^{V3} Relevance ratings would demonstrate incremental validity over the Presence ratings was not supported, consistent with previous research by Strub and colleagues (2014). This is possibly due to the ratings capturing similar information and becoming redundant when considered simultaneously. Indeed, Presence and Relevance ratings were significantly correlated and produced almost identical AUCs (Relevance ratings in fact producing a superior AUC to HCR-20^{V3} SRRs and Total scores). In regression analyses, both emerged as non-significant predictors in the final model and, in effect, cancelled each other out. Other studies have also made use of an interaction term between Presence and Relevance ratings (given the strength of association) and used this as a singular variable in regression analyses (Howe et al., 2016; Strub et al., 2014), which is conceptually more sound. Indeed, research has demonstrated that the interaction between Presence and Relevance ratings uniquely predicts SRRs above the sum of scores (Smith et al., 2014), and therefore the success of SRRs in yielding the largest AUC reflects the interaction between Presence and Relevance ratings. Interaction terms were not adopted in the current study. While the use of interaction terms holds utility for research purposes, clinically it is less applicable as clinicians would not generate interaction terms in practice and standard use of the HCR-20^{V3}.

Given the archival nature of the study, relevance of risk factors as assessed through file review may differ from relevance as assessed by clinicians who work directly with patients. Such familiarity may bring a richer understanding of the relevance of risk factors to an individual’s perpetration of violence. The performance of the Relevance ratings as a

novel concept in the current study has been promising, particularly given increased levels of subjectivity in judging relevance over simply identifying presence.

The finding of significant incremental validity afforded by the PCL-R to HCR-20^{V3} Total scores and SRRs was unexpected and contradicts some studies suggesting that the PCL-R does not add incremental predictive accuracy to the HCR-20 (Campbell, 2007; Douglas & Webster, 1999; Guy et al., 2010; Hogan & Olver, 2016). Given that the PCL-R captures a broad range of personality traits, behaviours and affective dispositions, it is reasonable that the PCL-R could add significantly to the model (Ogloff, 2006). As a measure of psychopathy, the PCL-R cannot substitute clinically for the risk-related outputs that the HCR-20^{V3} can provide (i.e. SRRs, risk scenarios and risk management plans). Conversely, the PCL-R provides a reliable and valid assessment of psychopathy that possibly exceeds the ability of clinicians to estimate or capture in H7 (Personality Disorders) unaided, without formal consideration of the individual traits. The current study supports continued inclusion of the PCL-R as part of a structured and comprehensive assessment of risk of violence.

Limitations and future research

The above findings should be considered in light of the study limitations, including retrospective design and small sample size. Reliance on file review to determine the relevance of risk factors may have been limiting and could have been enriched by the addition of interviews with the patients or staff. As recidivism was informed solely by police records, the base rate of violence recorded in this study was probably an underestimate of the true level of violence perpetrated (Douglas & Ogloff, 2003; Mulvey, Shaw, & Lidz, 1994). Notably, the LEAP database is also restricted to the Victorian state, and therefore offending in other Australian states or

territories would not have been identified during follow-up. Whilst a minority of the sample was female, inclusion of the Female Additional Manual (FAM; de Vogel, de Vries Robbé, et al., 2014) would have been ideal to inform gender-sensitive assessment.

Future research should endeavour towards a prospective design and confidential inquiry approach. Whilst the current adopted a longitudinal follow-up period, shorter follow-up periods are more useful to the clinical use of the tool in forecasting and decision making for the near future. Fixed follow-up periods would be helpful in reporting on validity outcomes for specific time points. Given that one third of the sample transferred directly to correctional settings, and data informing in-custody acts of violence were not accessed, the current procedure deviated significantly from standard use of the HCR-20 in using dynamic risk variables to predict short-term risk. Future research should endeavour to access data on inpatient and in-custody acts of violence and follow up participants' move through various institutional and community settings. Novel aspects of the HCR-20^{V3}, such as the Relevance ratings, alternative SRRs and risk scenarios, warrant further exploration, which is a far more complex task than evaluation of predictive validity.

Conclusion

The HCR-20 scheme for violence risk assessment has advanced considerably over the past two decades into its third iteration. This study evaluated select psychometric properties for the HCR-20^{V3} and represents one of the first evaluations of the HCR-20^{V3} in Australia. Results are promising, supporting the use of the HCR-20^{V3} within forensic psychiatric cohorts. However, findings suggest that evaluators may benefit from further training and guidance on unfamiliar elements of the HCR-20^{V3} to improve consistency and should consider including the PCL-R for assessments. As the evidence-base for the HCR-20^{V3} remains

in its infancy, further research evaluating novel elements of the tool is warranted.

Ethical standards

Declaration of conflicts of interest

Delene Brookstein has declared no conflicts of interest

Michael Daffern has declared no conflicts of interest

James Ogloff has declared no conflicts of interest

Rachel Campbell has declared no conflicts of interest

Chi Meng Chu 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.

References

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

Bonta, J., & Andrews, D.A. (2007). Risk-need-responsivity model for offender assessment and rehabilitation. *Rehabilitation, 6*, 1–22.

Campbell, R.E. (2007). *Antisocial personality disorder, psychopathy, and the assessment of risk for violence in an Australian mentally disordered population* (Doctoral Dissertation). Monash University, Australia.

Chu, C.M. (2010). The predictive accuracy of static and dynamic measures for assessing risk of inpatient aggression in a secure psychiatric hospital (Doctoral dissertation). Available from Monash University Library, Record No. 63184.

Crimes Act. (1958). (Vic)

de Vogel, V., de Vries Robbé, M., van Kalmthout, W., & Place, C. (2014). *FAM. Female Additional Manual: Additional guidelines to the HCR-20V3 for assessing risk for violence in women*. Utrecht, Netherlands: Van der Hoeven Kliniek.

Retrieved from <https://irp-cdn.multiscreensite.com/21b376df/MOBILE/pdf/fam+to+be+used+with+hcr-20+version+3+-+english+-+version+2014.pdf>.

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

DeLong, E.R., DeLong, D.M., & Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44*(3), 837–845. DOI: 10.2307/2531595.

Dickens, G.L., & O’Shea, L.E. (2017). Use of the HCR-20 for violence risk assessment: Views of clinicians working in a secure inpatient mental health setting. *The Journal of Forensic Practice, 19*(2), 130–138. doi: 10.1108/JFP-08-2016-0039

Dolan, M., & Doyle, M. (2000). Violence risk prediction. Clinical and actuarial measures and the role of the Psychopathy Checklist. *The British Journal of Psychiatry: The Journal of Mental Science, 177*, 303–311. doi:10.1192/bjp.177.4.303

Douglas, K.S., & 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.S., & Ogloff, J.R.P. (2003). Violence by psychiatric patients: The impact of archival measurement source on violence base rates and risk assessment accuracy. *Canadian journal of psychiatry. Revue canadienne de psychiatrie, 48*(11), 734–740. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.526.5413&rep=rep1&type=pdf>

Douglas, K.S., & Webster, C.D. (1999). The HCR-20 violence risk assessment scheme: Concurrent validity in a sample of incarcerated offenders. *Criminal Justice and Behavior, 26*(1), 3–19. doi:10.1177/0093854899026001001

Douglas, K.S., Hart, S.D., Webster, C.D., & Belfrage, H. (2013). *HCR-20 (Version 3): Assessing risk for violence – User Guide*. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Douglas, K.S., Hart, S.D., Webster, C.D., Belfrage, H., Guy, L.S., & Wilson, C.M. (2014). Historical-Clinical-Risk Management-20, Version 3 (HCR-20V3): Development

- and overview. *International Journal of Forensic Mental Health*, 13(2), 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: V3. *International Journal of Forensic Mental Health*, 13(2), 140–147. doi:10.1080/14999013.2014.906517
- Edens, J.F., Campbell, J.S., & Weir, J.M. (2007). Youth psychopathy and criminal recidivism: A meta-analysis of the Psychopathy Checklist measures. *Law and Human Behavior*, 31(1), 53–75. doi:10.1007/s10979-006-9019-y
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.
- Fleiss, J.L. (1981). *Statistical Methods for Rates and Proportions* (2nd ed.) New York: Wiley.
- Guy, L., Douglas, K.S., & Hendry, M.C. (2010). The role of psychopathic personality disorder in violence risk assessments using the HCR-20. *Journal of Personality Disorders*, 24(5), 551–580. doi:10.1521/pedi.2010.24.5.551
- Guy, L.S. (2008). Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey. Unpublished doctoral dissertation, Simon Fraser University, Burnaby, British Columbia, Canada. Retrieved from <http://summit.sfu.ca/item/9247>.
- Hanley, J.A., & McNeil, B.J. (1982). The meaning and use of the area under a Receiver Operating Characteristic (ROC) Curve. *Radiology*, 143(1), 29–36. Retrieved from: http://www.med.mcgill.ca/epidemiology/hanley/software/hanley_mcneil_radiology_82.pdf. doi:10.1148/radiology.143.1.7063747
- Hanley, J.A., & McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148(3), 839–843. Retrieved from: <http://www.med.mcgill.ca/epidemiology/hanley/reprints/> doi:10.1148/radiology.148.3.6878708Method_of_Comparing_1983.pdf.
- Hare, R.D. (1991). *The revised psychopathy checklist*. Toronto, Ontario, Canada: Multi-Healthy Systems.
- Hare, R.D. (2003). *Manual for the hare psychopathy checklist-revised*, 2nd ed. Toronto, Canada: Multi Health Systems.
- Hart, S.D., Cox, D.N., & Hare, R.D. (1995). *Manual for the psychopathy checklist: Screening version (PCL:SV)*. Toronto: Multi-Health Systems.
- Hogan, N.R., & Olver, M.E. (2016). Assessing risk for aggression in forensic psychiatric inpatients: An examination of five measures. *Law and Human Behavior*, 40(3), 233–243. doi:10.1037/lhb0000179
- Hogan, N.R., & Olver, M.E. (2019). Static and dynamic assessment of violence risk among discharged forensic patients. *Criminal Justice and Behavior*, 46(7), 923–938. doi:10.1177/0093854819846526
- Hopton, J., Cree, A., Thompson, S., Jones, R., & Jones, R. (2018). An evaluation of the quality of HCR-20 risk formulations: A comparison between HCR-20 Version 2 and HCR-20 Version 3. *International Journal of Forensic Mental Health*, 17(2), 195–201. doi:10.1080/14999013.2018.1460424
- 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
- Howell, D.C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Duxbury.
- Kötter, S., von Franqué, F., Bolzmacher, M., Eucker, S., Holzinger, B., & Müller-Isberner, R. (2014). The HCR-20^{V3} in Germany. *International Journal of Forensic Mental Health*, 13(2), 122–129. doi:10.1080/14999013.2014.911784
- Leistico, A.-M.R., Salekin, R.T., DeCoster, J., & Rogers, R. (2008). A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. *Law and Human Behavior*, 32(1), 28–45. doi:10.1007/s10979-007-9096-6
- Martinez-Camblor, P., Carleos, C., & Corral, N. (2013). General nonparametric ROC curve comparison. *Journal of the Korean Statistical Society*, 42(1), 71–81. doi:10.1016/j.jkss.2012.05.002
- Mastromanno, B., Brookstein, D.M., Ogloff, J.R., Campbell, R., Chu, C.M., & Daffern, M. (2018). Assessing change in dynamic risk factors in forensic psychiatric inpatients: Relationship with psychopathy and recidivism. *The Journal of Forensic Psychiatry & Psychology*, 29(2), 323–336. doi:10.1080/14789949.2017.1377277
- MedCalc Software Version 16.2.1. (2016). Ostend, Belgium. Retrieved from <https://www.medcalc.org/download.php>.
- Mental Health Act. (2014). (Vic)

- Monahan, J., Steadman, H.J., Silver, E., Appelbaum, P.S., Robbins, P.C., Mulvey, E.P., ... Banks, S. (2001). *Rethinking risk assessment: The MacArthur Study of Mental Disorder and Violence*. New York: Oxford University Press.
- Mossman, D. (2013). Evaluating risk assessments using receiver operating characteristic analysis: Rationale, advantages, insights, and limitations. *Behavioral Sciences & the Law*, 31(1), 23–39. doi:10.1002/bsl.2050
- Mulvey, E.P., Shaw, E., & Lidz, C.W. (1994). Why use multiple sources in research on patient violence in the community? *Criminal Behaviour and Mental Health*, 4(4), 253–258. doi:10.1002/cbm.1994.4.4.253
- Ogloff, J.R.P. (2006). The psychopathy/antisocial personality disorder conundrum. *The Australian and New Zealand Journal of Psychiatry*, 40(6–7), 519–528. doi:10.1080/j.1440-1614.2006.01834.x
- Penney, S.R., Marshall, L.A., & Simpson, A.I.F. (2016). The assessment of dynamic risk among forensic psychiatric patients transitioning to the community. *Law and Human Behavior*, 40(4), 374–386. doi:10.1037/lhb0000183
- Persson, M., Belfrage, H., Fredriksson, B., & Kristiansson, M. (2017). Violence during imprisonment, forensic psychiatric care, and probation: Correlations and predictive validity of the risk assessment instruments COVR, LSI-R, HCR-20^{V3}, and SAPROF. *International Journal of Forensic Mental Health*, 16(2), 117–129. doi:10.1080/14999013.2016.1266420
- Rice, M.E., & Harris, G.T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and Human Behavior*, 29(5), 615–620. doi:10.1007/s10979-005-6832-7
- Salekin, R.T., Rogers, R., & Sewell, K.W. (1996). A review and meta-analysis of the psychopathy checklist and psychopathy checklist-revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice*, 3(3), 203–215. doi:10.1111/j.1468-2850.1996.tb00071.x
- Singh, J., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and meta-regression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, 31(3), 499–513. doi:10.1016/j.cpr.2010.11.009
- Singh, J.P., Desmarais, S.L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., ... Otto, R.K. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 Countries. *International Journal of Forensic Mental Health*, 13(3), 193–206. doi:10.1080/14999013.2014.922141
- Singh, J.P., Yang, S., & Mulvey, E.P. (2015). Reporting guidance for violence risk assessment predictive validity studies: The RAGEE Statement. *Law and Human Behavior*, 39(1), 15–22. doi:10.1037/lhb0000090
- Smith, K.J., O'Rourke, S., & Macpherson, G. (2020). The predictive validity of the HCR20^{V3} within Scottish Forensic Inpatient Facilities: A closer look at key dynamic variables. *International Journal of Forensic Mental Health*, 19(1), 1–17. doi:10.1080/14999013.2019.1618999
- Smith, S., 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
- Strub, D.S., Douglas, K.S., & Nicholls, T.L. (2014). The validity of version 3 of the HCR-20 violence risk assessment scheme amongst offenders and civil psychiatric patients. *The International Journal of Forensic Mental Health*, 13(2), 148–159. doi:10.1080/14999013.2014.911785
- Victorian Institute of Forensic Mental Health. (2012). *Report of operations 2011–2012*. Fairfield, Victoria, Australia: Author. Retrieved from <http://www.forensicare.vic.gov.au>
- Victorian Institute of Forensic Mental Health. (2015). *Quality of care report*. Fairfield, Victoria, Australia: Author. Retrieved from <http://www.forensicare.vic.gov.au/assets/pubs/QualityofCareReport20142015.pdf>
- Webster, C., Douglas, K., Eaves, D., & Hart, S. (1997). *HCR-20: Assessing risk for violence (version 2)*. Burnaby, BC: Simon Fraser University, Mental Health, Law and Policy Institute.