ARTICLE

Evaluation of a modified procedure for rating the presence of HCR-20 V3 risk factors

Philip Minoudis[1] | Bardana Singh[2] | Chrissie Lane[2] |
Marvin Iroegbu[1] | Caitriona Higgins[2] | Clare Wellington[1] |
Stephen D. Hart[3] | Kevin S. Douglas[3] | Mark Freestone[4]

[1]London Pathways Partnership and Oxleas NHS Foundation Trust, Bexley, UK

[2]London Pathways Partnership and East London NHS Foundation Trust, London, UK

[3]Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada

[4]Wolfson Institute of Population Health, Queen Mary University of London, London, UK

Correspondence
Philip Minoudis, London Pathways Partnership and Oxleas NHS Foundation Trust, Bexley, UK.
Email: philip.minoudis@nhs.net

Abstract

Introduction: We evaluated a modified procedure for making more detailed ratings of the presence of risk factors for violence using the HCR-20 V3 in a field study.

Method: Participants were 225 offenders recruited from correctional and forensic mental health institutions who were assessed by trained clinicians. We examined the measurement precision and predictive validity of ratings made using the modified procedure compared to ratings made by the same clinician using the standard procedure.

Results: Ratings made using the modified procedure had equivalent or superior internal consistency to the original rating procedure and had equivalent predictive accuracy (validity) with respect to future violence according to logistic regression analyses.

Discussion and conclusion: These findings suggest that the modified procedure may increase the sensitivity of Presence ratings of the HCR-20 V3 risk factors.

KEYWORDS
dynamic risk factors, HCR-20 V3, risk assessment, violence

INTRODUCTION

The past 30 years have seen important advances in the development of structured decision support aids for violence risk assessment. If these decision support aids improve the reliability and validity of professional judgements about violence risk, they have the potential to help decrease the rate at which or length of time for which people are detained in correctional or healthcare institutions, improve the appropriateness of rehabilitative services people receive in institutional or community settings and enhance public safety. If not, they will contribute to the continuation—or even the worsening—of the economic and human costs of violence.

One popular decision support aid is the Historical Clinical Risk Management-20 or HCR-20. The HCR-20 is a set of structured professional judgement (SPJ) guidelines—evidence-based practice parameters developed from a systematic review of the scientific and professional literature. To date, three versions of the HCR-20 have been published: the original version, now referred to as HCR-20 V1 (Webster et al., 1995); the second version, HCR-20 V2 (Webster et al., 1997) and the third version, HCR-20 V3 (Douglas et al., 2013). We refer readers elsewhere to discussions of the SPJ approach (Hart et al., 2016; Hart & Douglas, 2019, 2023) and for additional information concerning the administration of the three versions of the HCR-20 (Douglas & Shaffer, 2020). Briefly, one important step in the administration of all three versions is consideration of the state of 20 basic risk factors, as well as any additional case-specific risk factors. The basic risk factors are divided into three domains. The historical domain comprises 10 factors that reflect specific aspects of the person's entire history of psychosocial functioning. The clinical (C factors) domain comprises five factors that reflect aspects of the person's recent and current psychosocial functioning. Finally, the risk management (R factors) domain comprises five factors that reflect aspects of the person's anticipated psychosocial functioning in light of their plans for the future. Independent evaluations of the three versions of the HCR-20 have been conducted in more than three dozen countries, and an international survey of 2000 clinicians indicated the HCR-20 V2 was the most commonly used decision support aid for violence risk assessment (Singh et al., 2014). Research indicates that judgements of violence risk made using the three versions of the HCR-20 have good interrater reliability as well as predictive validity with respect to violence in institutional and community settings that is moderate in absolute terms and, in relative terms, equal or superior to that of other decision support tools (for a recent review, see Douglas & Shaffer, 2020).

The SPJ approach is intended to assist comprehensive and individualized assessments of violence risk. To that end, SPJ guidelines such as the HCR-20 are designed to be sensitive to dynamic aspects of risk. Assessment of change in risk over time is a critical issue, as many legal and clinical decisions are based on the risks *currently* posed by people and thus cannot be determined solely from a review of factors that are static or historical in nature (e.g. Department of Health, 2007). While some prison and probation predictive risk assessments may include dynamic factors, these are not designed to assess change in clinical risk factors that are targeted in interventions. For example, in England and Wales, the Offender Personality Disorder Pathway—operating out of more than 60 correctional and forensic mental health offices in institutional and community settings—uses structured decision support aids to help evaluate both the violence risk posed by offenders and the effectiveness of services delivered to them, but the value of these decision support aids depends heavily on their ability to assess change in risk.

But what is the evidence that the HCR-20 can be used to reliably assess changes in risk? Several studies addressed this question by examining changes over time in ratings of the state of the basic risk factors included in the HCR-20 V2, referred to as Presence ratings.[1] As the risk factors in the Historical domain are expected to change little or slowly over time, these studies all focused on changes in the Presence ratings for the risk factors in the Clinical and Risk Management domains. Per the standard procedure for administering the HCR-20 V2, Presence ratings were made using a 3-point ordinal response format: *Present*, *Possibly or partially present* or *Not present*. For research purposes, these ordinal responses are typically recoded numerically as *Present* = 2, *Possibly or partially present* = 1 and *Not present* = 0 and sometimes summed to yield total and domain composite scores.[2] Two studies of forensic inpatients

---

[1]Changes in risk could also be indexed by judgements of overall risk, sometimes referred to as summary risk ratings or care prioritization ratings, using all three versions of the HCR-20. In addition, a new step was added to the administration procedure of the HCR-20 V3 for rating the Relevance of risk factors, that is, their putative causal relevance with respect to the perpetration of violence by the person or to interference with management of the person's risk. In this paper, we focus solely on the potential use of Presence ratings in the evaluation of changes in risk.

[2]The purpose of numerically recoding Presence ratings was to facilitate the evaluation of the reliability and validity of risk factors, both individually and as an ensemble. As these ratings have been treated as continuous rather than categorical or ordinal in nature in past research, we do the same in this paper, based on the pragmatic view that "everything is linear to a first approximation" (Pasta, 2009, p. 31). We acknowledge, though, that this assumption can and should be subjected to empirical evaluation.

observed a significant decrease in Presence ratings over time (Belfrage & Douglas, 2002; De Vries Robbé et al., 2015), presumably as a result of treatment; a third study did not report a decrease, although the authors speculated that this could have been an artefact of the study method (K. Wilson et al., 2014). All these studies used the HCR-20 V2. Three other studies, two of forensic inpatients (C. Wilson et al., 2013; Hogan & Olver, 2016) and one of a mixed sample of civil and forensic outpatients (Michel et al., 2013), reported that the Presence ratings of C and R factors change substantially over time and, furthermore, increases or decreases in Presence ratings were also significantly associated with increases or decreases in the actual likelihood of subsequent violence. Two of these studies used the HCR-20 V2 and one used the HCR-20 V3.

Although these results support the potential utility of Presence ratings for assessing change in risk, a potential limitation of the HCR-20 in this respect is its reliance on a 3-point response format for those ratings. Indeed, this possibility was raised in the evaluation of another decision support aid for violence risk assessment modelled on the HCR-20 (Verschueren et al., 2023). It is possible that small yet consequential changes in the state of risk factors are observable to evaluators but not detectable by this relatively crude response format. Research stretching back many decades has found that ratings made using simple 2- or 3-point response formats are typically inferior with respect to reliability and validity compared to ratings made using more complex response formats (e.g., Cox, 1980, Chang, 1994; Colman et al., 1997; Hancock & Klockars, 1991; Loken et al., 1987). Typically, response formats of 5 or 7 points are considered optimal, with response formats exceeding 7 points yielding diminishing returns, increasing the complexity of ratings with little or no incremental reliability or validity (Ghiselli, 1955; Green & Rao, 1970; Preston & Colman, 2000; Symonds, 1924).

The present study

We developed a 7-point response format for HCR-20 V3 Presence ratings. We then conducted field research to evaluate Presence ratings made using the standard 3-point response format versus those made using the novel 7-point response format in a mixed sample of correctional offenders and forensic mental health patients. In this paper, we report the results of analyses that addressed four research questions:

1. Do HCR-20 V3 Presence ratings for individual risk factors and for total and domain composites made using the novel 7-point response format have acceptable interrater reliability?
2. Do HCR-20 V3 Presence ratings for individual risk factors and for total and domain composites made using the novel 7-point response format have a strong association with those made using the standard 3-point response format?
3. Do HCR-20 V3 Presence ratings for individual risk factors within domains made using the novel 7-point response format have internal consistency reliability[3] equal to or superior to that of ratings made using the standard 3-point response format?
4. Do HCR-20 V3 Presence ratings for individual risk factors and for total and domain composites made using the novel 7-point response format have criterion-related validity with respect to recent violence equal to or superior to that of ratings made using the standard 3-point response format?

Answers in the affirmative to all four questions would support (more correctly, fail to disconfirm) the potential utility of the novel 7-point response format for Presence ratings to assess the risk factors; an answer in the negative to one or more questions would disconfirm their potential utility. This is the

---

[3]As a risk assessment tool, the HCR-20 is based on a formative rather than reflective model of assessment. Thus, Cronbach's here is simply an index of the reliability of the composite Presence ratings for risk factors within domains as an ensemble of predictors. The Presence ratings are not, and should not be conceptualized or interpreted as, indicators of a latent trait; and Cronbach's is not, and should not be conceptualized or interpreted as, an index of the unidimensionality of a latent trait.

first stage in developing a longer response format which may improve the ability to assess change over time, although this element was beyond the scope of this paper.

METHOD

This study's design and its analysis were not preregistered.

Participants

The study was conducted in the United Kingdom. An independent researcher from His Majesty's Prison and Probation Service (HMPPS) invited participation from Offender Personality Disorder (OPD) services in HMPPS. Five prison OPD services initially expressed interest. National Health Service (NHS) medium secure services forming part of a consortium of Trusts that provide OPD services to London, the London Pathways Partnership (LPP), were invited to participate and accepted. An additional Mental Health NHS Foundation Trust, which had collaborated on initial planning of the research, was also invited to take part. One prison and one mental health site withdrew from the study due to research time restrictions, leaving four prison (HMPPS) sites and four forensic mental health (NHS) sites that participated fully in the research.

The study received ethical approval to carry out the research in the HMPPS and NHS sites, respectively. The Research Ethics Committee stipulated the importance of unrestricted access to HCR-20 V3 forms and recommended, under section 251 of the *NHS Act* (2006), the collection of data without consent provided it could not be used to identify individual service users. Although permission to collect this data without direct participant consent was granted by the UK Confidentiality Advisory Group (CAG), we decided to inform potential participants about the study and their right to withdraw. However, no participants chose to withdraw. Sites were encouraged to obtain consent from potential participants directly via posters and direct contact with service users. This obviated the need for study researchers to recruit participants, although they were available to answer questions about the study at the request of service users. To protect the privacy of participants, data that could be used to identify individual service users were not collected.

A total of 230 service users from the eight sites were recruited: 78 from HMPPS sites and 152 from NHS sites. Of these, 5 service users were excluded because they were missing Presence ratings for more than 5 of 20 HCR-20 V3 risk factors overall or for more than 2 risk factors in any domain. Thus, the final sample comprised 225 participants. All participants had received clinical diagnoses of personality disorder or clinically significant symptoms of personality disorder prior to receiving OPD services. Of these, 129 participants had a primary diagnosis of personality disorder (PD) and 96 participants (all from NHS sites) had a primary diagnosis of a major mental disorder (e.g. a severe psychotic disorder or a severe neurodevelopmental, neurocognitive, or mood disorder). All participants were male.

Because we did not collect data that could be used to identify individual service users, we are unable to present detail concerning the demographic characteristics of participants. Instead, the demographic characteristics of the service users at the HMPPS and NHS sites from which participants were recruited are summarized in Tables 1a and 1b.

Procedure

Development of the novel 7-point response format for Presence ratings

The research team, which includes two authors of the HCR-20 V3, developed a 7-point response format for Presence ratings, which ranged from *Not present* (0) to *Definitely present* (6). We also developed written

TABLE 1A Demographics of prison sites.

Demographic variables	Prison #1*		Prison #2*		Prison #3*		Prison #4*	
	n	%	*n*	%	*n*	%	*n*	%
Age (years)								
18–29	310	32.1	70	22.2	141	16.7	194	26.9
30–39	310	32.1	113	35.9	234	27.8	236	32.7
40–49	186	19.3	69	21.9	164	19.5	131	18.2
50–59	106	11.0	46	14.6	177	21.0	109	15.1
60–69	32	3.3	13	4.1	89	10.6	35	4.9
70+	21	2.2	4	1.3	36	4.3	16	2.2
Max age	86	-	74	-	82	-	85	-
Ethnicity								
Asian or Asian British	65	6.7	32	10.1	46	5.8	90	12.5
Black or Black British	265	27.5	87	27.7	71	9	211	29.3
Mixed	54	5.6	23	7.2	28	3.5	37	5.2
White	544	56.5	164	52	50	6.1	364	50.5
Other Ethnic Group	24	2.5	6	1.9	3	0.4	16	2.2
Length of stay (sentenced prisoners only)								
Less than 1 month	38	3.9	5	1.6	13	1.5	104	14.4
1–3 months	78	8.1	14	4.4	39	4.6	170	23.6
3–6 months	97	10.1	15	4.8	100	11.9	212	19.4
6–12 months	113	11.7	37	11.7	154	18.3	172	23.9
1–2 years	209	21.7	28	8.9	191	22.7	49	6.8
2–4 years	313	32.4	97	30.8	66	7.8	11	1.5
4+ years	117	12.1	114	36.2	272	32.3	2	0.3
Security category								
Category A high risk	-	-	8	2.5	9	1.1	-	-
Category A	-	-	127	40.3	235	27.9	-	-
Provisional Category A	-	-	1	0.3	7	0.8	-	-
Category B	736	76.3	178	56.5	586	69.8	-	-
Category C	222	23.0	-	-	3	0.4	670	92.9
Category D	7	0.7	-	-	-	-	49	6.8

Note: *n* = total number of individuals.

*Prisons included OPD units at HMPs Swaleside, Frankland, Brixton and Whitemoor, not identified against demographics to preserve anonymity.

instructions that included a definition for each of the 7 anchor points. Further detail concerning the novel 7-point response format is available in the Supplemental Material S1.

Recruitment and training of clinicians

The research team recruited clinicians from various sites. The clinicians were qualified psychologists or psychiatrists responsible for the care of offenders or patients who had consented to participate in the study.

TABLE 1B Demographics of NHS sites.

Demographic variables	MSU #1		MSU #2		MSU #3		MSU #4	
	n	%	*n*	%	*n*	%	*n*	%
Age (years)								
18–20							3	1.5
21–29	65 (19–35)	35.2	0	0	53	28.3	35	16.9
30–39	119 (36–79)	64.7	2	14.3	61	32.6	67	32.4
40–49			6	42.9	43	23.0	43	20.8
50–79			6	42.8	30	16.0	59	28.5
Ethnicity								
Asian	53	59.4	3	25	114	61.3	87	42.0
Mixed	7	7.7	1	8	7	3.76	17	8.2
White	0	0	8	67	52	28.0	83	40.1
Other Ethnic Group	5	5.6	0	0	13	7.0	15	7.25
Length of stay (days)								
Average	116.3	-	187	-	1156.5	-	1374	-

Note: *n* = total number of individuals.

*NHS sites included MSUs River House (SLaM), John Howard Centre, Bracton Centre and Camlet Lodge, not identified against demographics to preserve anonymity.

All the clinicians completed training and supervised practice in the usual administration of the HCR-20 V3, which included the use of the standard 3-point response format for making Presence ratings. Some had completed training and supervised practice prior to their recruitment into the study, and others completed training and supervised practice after their recruitment but prior to administering the HCR-20 V3 for the purpose of the study.

All the clinicians also received specialized training in the use of the novel 7-point response format for making Presence ratings. Some completed this training in person, and others completed it online in the form of a narrated PowerPoint presentation. Further detail concerning the training program for the use of the novel 7-point response format for making Presence ratings is available in the Supplemental Material S1.

Presence ratings

Clinicians completed the HCR-20 V3 according to the usual administration procedures as part of routine clinical practice. The HCR-20 V3 comprises 20 risk factors, including 10 historical, 5 clinical and 5 future-based risk factors. The administration was based on all available clinical data: interview, observation and review of information in the data systems of the prison sites (pNOMIS,) or the forensic mental health sites (Rio or ePatientJourney). As part of this, clinicians made Presence ratings for individual risk factors using the standard 3-point response format. For research purposes, they also made Presence ratings for individual risk factors using the novel 7-point response format.

Presence ratings made by clinicians and researchers were collated after being stripped of all identifying information. Presence ratings were then recoded from ordinal categories into numeric scores. For the standard 3-point response format, the Presence scores were 0 (*Not present*), 1 (*Possibly/partially present*) or 2 (*Present*); and for the novel 7-point response format, the Presence scores ranged from 0 (*Not present*) to 6 (*Definitely present*). Presence scores could not be calculated for a relatively small number of individual risk factors (no more than 5 overall and no more than 2 in any domain) and in a relatively small number of cases (37 of 225, or 16%, had at least one missing data point); this was because Presence ratings were

TABLE 2 Baseline descriptive statistics for 3- and 7-point scales for Presence ratings, raw data (*N* = 230).

Item/scale	*n*	3-point scale		7-point scale	
		M	SD	M	SD
H1	225	1.95	0.25	4.53	1.45
H2	224	1.56	0.73	3.77	1.95
H3	223	1.79	0.48	4.67	1.48
H4	223	1.57	0.67	3.88	1.71
H5	225	1.64	0.72	4.12	2.09
H6	223	1.50	0.79	4.10	2.02
H7	210	1.37	0.82	3.77	2.40
H8	219	1.67	0.61	4.14	2.03
H9	220	1.40	0.76	3.43	1.87
H10	223	1.66	0.58	4.09	1.54
H total*	199	16.71	2.60	43.68	9.21
C1	225	1.28	0.70	3.69	1.51
C2	220	0.86	0.79	2.05	1.86
C3	224	1.11	0.82	2.85	2.07
C4	223	1.23	0.79	3.21	2.06
C5	225	1.18	0.65	3.19	1.74
C total*	199	5.79	2.46	15.82	6.38
R1	223	0.98	0.75	2.82	1.84
R2	224	1.17	0.74	3.04	1.88
R3	225	1.35	0.65	3.83	1.50
R4	224	1.33	0.63	3.67	1.51
R5	224	1.67	0.51	3.35	1.19
R total	199	6.39	2.27	17.80	5.70
HCR20 total*	199	28.74	5.58	77.31	17.46

*Summary statistics for complete cases only.

sometimes omitted by clinicians due to an absence of case history information. Missing data points for individual risk factors were imputed by means of multiple imputation by chained equations (Azur et al., 2011) using the *mice* package in R (Buuren & Groothis-Oudshoorn, 2010). We then formed composite total and domain Presence scores by summing scores for individual risk factors. Historical scores were the sum of scores for H1 to H10, ranging from 0 to 20 for 3-point ratings and from 0 to 60 for 7-point ratings. Clinical scores were the sum of scores for C1 to C5, ranging from 0 to 10 for 3-point ratings and from 0 to 30 for 7-point ratings. Risk Management scores were the sum of scores for R1 to R5, ranging from 0 to 10 for 3-point ratings and from 0 to 30 for 7-point ratings. Total scores were the sum of Historical, Clinical, and Risk Management scores, ranging from 0 to 40 for 3-point ratings and from 0 to 120 for 7-point ratings. The distributions of numerical Presence scores for individual risk factors and composites made by clinicians using 3- and 7-point response formats are presented in Table 2.

As usual, practice in the sites was to have the HCR-20 V3 completed by a single responsible clinician (albeit often in consultation with other professionals involved in the delivery of services to offenders or patients); we were unable to evaluate the interrater reliability of all judgements made by clinicians using the HCR-20 V3. We did not view this as a significant limitation, as previous research—including research conducted at some of the sites from which participants were recruited—indicates that Presence ratings for individual risk factors made using the standard 3-point response format typically have adequate (i.e. fair to good) interrater reliability (see Douglas & Schafer, 2020).

We did, however, evaluate the interrater reliability of Presence ratings made using the novel 7-point response format. To this end, we selected a random subsample of 23 participants. For each participant, at least one experienced researcher chosen from a pool of four reviewed the information available on the prison or forensic mental health data systems and made Presence ratings for individual risk factors using the novel 7-point response format. The researcher was blinded to the Presence ratings made by clinicians in either response format. The results were analysed using two-way random effects intraclass correlation coefficients (ICCs) as the reliability ratings were made by a pool of assessors.

Institutional violence in 6 months prior to assessment

After the Presence ratings were made, researchers coded proximal violence in the 6 months prior to assessment from the data systems of the prison sites (pNOMIS) or the forensic mental health sites (Rio or ePatientJourney). We focused on violence perpetrated by participants during this time-frame. Researchers coded the presence of minor or major violence based on the definitions from the MacArthur Violence Risk Assessment (Monahan et al., 2001). For the purpose of the present study, we collapsed these codes into a single dichotomous variable scored 0 (*no violence*) versus 1 (*at least one incident or minor or major violence*). Overall, 66 of 230 participants (28.7%) had at least one incident of minor or major violence.

Data analyses

All analyses were conducted using R for Windows x 64 Version 4.2.3 (R Core Team, 2013). The R markdown file is available in the Supplemental Material S1.

Transparency and openness

All data, analysis code and research materials are available upon request from the first author.

RESULTS

Reliability

Interrater agreement

ICC (2, 1) (single measures) and ICC (2, k) (average measures) for individual risk factors and composites are presented in Table 3. For the Presence ratings, four ICC (2, 1) ratings were in the 'good' (Fleiss et al., 1981) range (0.75–0.9), and 16 were in the 'excellent' (Fleiss et al., 1981) range (>0.9), with 8 of these 16 being in the 'almost perfect' (Landis & Koch, 1977) range. Values of ICC (2, k), expectedly, were higher for all items, ranging from 0.82 to 0.96.

Scale level analyses of Presence

Findings presented in Table 4 are based on sums of Presence ratings across the items for History, Clinical and Risk, respectively, along with the total items in the scale (HCR-20 total). For ICC (2, 1) Presence ratings, two ratings were in the 'slight' range (Landis & Koch, 1977), one in 'fair' and one in

TABLE 3 Interrater reliability of Presence ratings of individual HCR-20 7-point Likert style.

HCR20 scale items	Presence ratings	
Historical scale (history of problems with…)	ICC (2, 1)	ICC (2, k)
H1. Violence	0.80 (0.58–0.91)	0.89 (0.74–0.95)
H2. Other Antisocial Behaviour	0.91 (0.81–00.96)	.96 (0.89–0.98)
H3. Relationships	0.86 (0.60–0.95)	0.92 (0.75–0.97)
H4. Employment	0.91 (0.81–00.96)	.96 (0.90–0.98)
H5. Substance Use	0.64 (0.32–0.83)	0.78 (0.49–0.91)
H6. Major Mental Disorder	0.85 (0.67–0.93)	0.92 (0.80–0.97)
H7. Personality Disorder	0.77 (0.53–0.89)	0.87 (0.69–0.94)
H8. Traumatic Experiences	0.72 (0.45–0.87)	0.83 (0.62–0.93)
H9. Violent Attitudes	0.90 (0.77–0.95)	0.94 (0.87–0.98)
H10. Treatment or Supervision Response	0.69 (0.40–0.86)	0.82 (0.57–0.92)
Clinical Scale (Recent Problems with…)		
C1. Insight	0.70 (0.23–0.88)	0.82 (0.38–0.94)
C2. Violent Ideation	0.80 (0.59–0.91)	0.89 (0.74–0.95)
C3. Symptoms of Major Mental Disorder	0.80 (0.59–0.91)	0.89 (0.74–0.95)
C4. Instability	0.80 (0.54–0.92)	0.89 (0.70–0.96)
C5. Treatment or Supervision Response	0.91 (0.80–00.96)	.96 (0.89–0.98)
Risk Management Scale (Future Problems with…)		
R1. Professional Services and Plans	0.93 (0.85–0.97)	0.96 (0.92–0.99)
R2. Living Situation	0.93 (0.84–97)	0.96 (0.92–0.99)
R3. Personal Support	0.76 (0.52–0.89)	0.87 (0.68–0.94)
R4. Treatment or Supervision Response	0.88 (0.74–0.95)	0.94 (0.85–0.97)
R5. Stress or Coping	0.79 (0.58–0.91)	0.89 (0.73–0.95)

Note: 95% confidence intervals are in parentheses.

TABLE 4 Interrater reliability of HCR20 7-point Likert scale, P.

Scale/judgement	Presence	
	ICC1	ICC2
H	0.93 (0.84–0.97)	0.96 (0.91–0.98)
C	0.73 (0.46–0.87)	0.84 (0.63–0.93)
R	0.97 (0.93–0.99)	0.98 (0.96–0.99)
HCR-20 total	(0.999 (1.0 to 2.d.p) (0.95–1.0))	(0.999 (1.0 to 2.d.p) (0.97–1.0))

Note: 95% confidence intervals in parentheses.

the 'moderate' range (Fleiss et al., 1981; Landis & Koch, 1977). Values of ICC (2, k), expectedly, were higher, ranging between 0.83 and 0.91.

Findings in Table 5 duplicate those in Table 4 but exclude items on Mental Illness (H6 and C3) to assess whether ICCs would increase. For ICC (2, 1) Presence ratings, one was in the 'fair' range (Landis & Koch, 1977), with the rest in the 'moderate' range (Fleiss et al., 1981; Landis & Koch, 1977), ranging between 0.43 and 0.50. Values for ICC (2, k) again were higher, ranging between 0.89 and 0.94.

As shown by the results, excluding these two items in the analyses yielded higher scores.

TABLE 5 Interrater reliability of HCR-20 7-point Likert style excluding Mental Illness items, Presence ratings.

	Presence	
Scale/judgement	ICC (2, 1)	ICC (2, k)
H (Excluding H6)	0.94 (0.87–0.98)	0.97 (0.93–0.99)
C (Excluding C3)	0.73 (0.47–0.88)	0.85 (0.64–0.93)
R	97 (0.93–0.99)	0.98 (0.96–0.99)
HCR-20 total (excluding H6 and C3)	(0.999 (1.0–2.d.p) (0.94–1.0))	(0.999 (1.0–2.d.p) (0.97–1.0))

Note: 95% confidence intervals in parentheses.

TABLE 6 HCR20v3 Scale reliability, presence items and imputed data ($n = 225$).

Scale	Cronbach α, 3-point	Cronbach α, 7-point
Historical	0.54	0.72*
Clinical	0.66	0.69
Risk	0.68	0.75
Total	0.76	0.83

*Significant difference, $p < .05$.

Internal consistency

Cronbach's alpha (α) for the total and domain composites is presented in Table 6. They ranged between 0.54 and 0.83, *Mdn* = .71. Two patterns are evident. First, internal consistency was higher for the total composite than for the domain composites. This is expected, as internal consistency is dependent on the number of items in a composite. Second, and crucially, there was no evidence that internal consistency was lower for 7-point ratings than for 3-point ratings. Indeed, the internal consistency was slightly higher for 7-point ratings than for 3-point ratings for all the composites. For this reason, it was unnecessary to conduct statistical tests to conclude that there was no evidence the internal consistency of composites based on 7-point ratings was significantly lower than that of composites based on 3-point ratings.

Predictive validity

H domain

Logistic regression modelling showed few differences between the 3-point and 7-point scoring models for the H items, with the 7-point model more sensitive for two items. In the 3-point scoring model, only H2 (Other Antisocial Behaviour) was significantly positively associated with proximal violence, OR = 1.67, 95% CI [1.07, 2.28], $p < .001$. When adjusted for all other factors, none remained significantly associated with the outcome. Using the 7-point model, H2 was again significantly associated with violence, OR = 1.28, 95% CI [1.02, 1.53], $p < .001$, as were two other factors: H1 (Prior Violence), OR = 1.43, 95% CI [1.16, 1.69], $p < .001$; and H3 (Relationships), OR = 1.28, 95% CI [1.02, 1.53], $p < .001$. Item H4 (Employment) showed an OR = 1.20, 95% CI [0.99, 1.40], $p = .05$. However, in the adjusted model, only H2 remained significantly associated with proximal violence, OR = 1.31, 95% CI [1.04, 1.57], $p < .001$. The results of these analyses are summarized in Tables 7 and 8.

TABLE 7 Logistic regression models for proximal violent acts, 3-point scoring.

	Main effects OR (95% CI)										Adjusted AOR (CIs)
	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	
H1	1.714 (0.500, 3.929)										1.470 (0.841, 3.780)
H2		1.673*** (1.070, 2.275)									1.510 (0.787, 2.233)
H3			1.642 (0.802, 2.483)								1.488 (0.586, 2.390)
H4				1.203 (0.661, 1.746)							0.952 (0.317, 1.587)
H5					1.145 (0.606, 1.684)						1.072 (0.494, 1.651)
H6						0.906 (0.535, 1.276)					0.943 (0.530, 1.356)
H7							1.112 (0.691, 1.533)				0.934 (0.416, 1.452)
H8								1.023 (0.455, 1.591)			0.888 (0.239, 1.537)
H9									1.236 (0.793, 1.680)		1.090 (0.570, 1.610)
H10										1.295 (0.684, 1.906)	1.138 (0.473, 1.804)
Constant	0.146 (−4.248, 4.540)	0.174 (−0.936, 1.285)	0.169 (−1.439, 1.776)	0.312 (−0.640, 1.264)	0.334 (−0.661, 1.329)	0.480* (−0.083, 1.043)	0.360 (−0.356, 1.077)	0.407 (−0.633, 1.446)	0.306 (−0.451, 1.063)	0.269 (−0.854, 1.392)	0.046 (−5.042, 5.134)
Obs	185	185	185	185	185	185	185	185	185	185	185

Note: *$p < .05$; **$p < .01$; ***$p < .001$.

TABLE 8 Logistic regression models for proximal violent acts, 7-point scoring.

	Main effects OR (95% CI)										Adjusted AOR (CIs)
H1	1.425*** (1.158, 1.692)										1.254 (0.911, 1.597)
H2		1.348*** (1.153, 1.543)									1.309*** (1.044, 1.573)
H3			1.276*** (1.019, 1.534)								1.119 (0.811, 1.426)
H4				1.199* (0.999, 1.398)							1.026 (0.774, 1.277)
H5					1.075 (0.912, 1.238)						0.919 (0.724, 1.113)
H6						0.948 (0.814, 1.082)					0.984 (0.823, 1.146)
H7							1.053 (0.915, 1.192)				0.917 (0.714, 1.121)
H8								1.164 (0.987, 1.340)			1.139 (0.920, 1.359)
H9									1.136 (0.977, 1.296)		0.931 (0.689, 1.174)
H10										1.182 (0.977, 1.386)	0.984 (0.734, 1.235)
Constant	0.075 (−1.322, 1.471)	0.114 (−0.851, 1.079)	0.127 (−1.228, 1.482)	0.203 (−0.703, 1.110)	0.315 (−0.478, 1.107)	0.526* (−0.029, 1.082)	0.346 (−0.341, 1.032)	0.215 (−0.675, 1.105)	0.260 (−0.460, 0.980)	0.209 (−0.753, 1.171)	0.038 (−2.065, 2.141)
Obs	208	208	208	208	208	208	208	208	208	208	208

Note: *p < .05; **p < .01; ***p < .001.

C domain

For the C domain, univariate analyses for both the 3-point and 7-point rating scales indicated that three items were significantly related to proximal violence: C2 (Violent Ideation or Intent), C4 (Instability) and C5 (Treatment or Supervision Response). The only difference between the scoring models was that in the 3-point scoring fully adjusted model, C4 (OR = 1.95, 95% CI 1.30–2.60, $p < .001$) and C5 (OR = 3.43, 95% CI 2.74–4.12, $p < .001$) both remained significantly associated with proximal violence, whereas in the 7-point model only C5 remained significant (OR = 1.29, 95% CI = 1.05, 1.53, $p < .001$), confidence intervals for C2 and C4 were both highly asymmetric, suggesting that with increased statistical power these items might have been significant in the 7-point adjusted model (see Tables 9 and 10).

R scales

For the R scales, the 3-point and 7-point scorings revealed different patterns of association between scale items and proximal violence. For the 3-point model, R3 (Personal Support) and R4 (Treatment or Supervision Response) were significantly related to proximal violence in the main effects analysis, but only R4 remained significant when adjusting for all other items (OR = 2.43, 95% CI = 1.80–3.06, $p < .001$). With the 7-point scoring model, R2 (Living Situation) and R4 were significantly associated with proximal violence as main effects, with again only R4 remaining a significant predictor after adjustment (OR = 1.45, 95% CI = 1.18–1.72, $p < .001$). We noted highly asymmetric confidence intervals for R2 (0.97–1.36) and also for R5 (0.50–1.08; also in the 3-point model 0.13–1.56) with the latter potentially suggestive of negative predictive value in a larger sample (see Tables 11 and 12).

DISCUSSION

This study asked three research questions to evaluate whether a novel 7-point response format on the HCR20 V3 performed as well as the standard 3-point response format in terms of interrater reliability, internal consistency and criterion-related validity. A fourth research question examined whether the 7-point response format had a strong association with the 3-point format in relation to Presence ratings for individual risk factors. The 3-point and 7-point response formats of the HCR-20 V3 performed similarly across these various tests. This supports, or at least does not disconfirm, that the novel 7-point response format for Presence ratings has comparable validity in practice to the widely used 3-point response format. The utility of a 7-point scale over a 3-point scale is to improve the sensitivity to assess change over time in the state of risk factors, although assessing change over time was beyond the scope of this project. This research supports the potential to improve assessment of change over time in relation to Presence of risk factors by using a longer 7-point response format on the HCR20 V3.

The interrater agreement on Presence ratings on the 7-point response format was rated in the 'good' to 'almost perfect' range, with 80% of items falling in the higher 'excellent' to 'almost perfect' range. This compared favourably to the standard 3-point response format, which typically has 'fair' to 'good' interrater reliability (see Douglas & Schafer, 2020). The number of cases used to test interrater reliability was at the lower end of the recommended sampling ratio, that is, 5%. This may have provided a more favourable impression of the strength of interrater reliability. Nevertheless, this finding supports the 7-point response format as having equal or superior interrater reliability compared to the 3-point format.

The 7-point response format performed at least as well or better than the 3-point format on a test of internal consistency using classical test theory. For the composite Historical scale, the 7-point response format had significantly higher internal consistency than the 3-point format, and the remaining Clinical, Risk and Total scores were all higher on the 7-point format, albeit not to a significant degree.

TABLE 9 Logistic regression models for proximal violent acts, 3-point scoring.

Main effects OR (95% CI)						Fully adjusted AOR (CI)
C1	0.892 (0.426, 1.357)					0.501 (−0.100, 1.102)
C2		1.843*** (1.433, 2.253)				1.256 (0.738, 1.774)
C3			0.886 (0.506, 1.267)			0.648 (0.197, 1.099)
C4				2.549*** (2.029, 3.068)		1.950*** (1.303, 2.597)
C5					2.897*** (2.348, 3.445)	3.431*** (2.738, 4.123)
Constant	0.493 (−0.199, 1.186)	0.204 (−0.410, 0.819)	0.475* (−0.001, 0.950)	0.110 (−0.756, 0.976)	0.114 (−0.680, 0.907)	0.101 (−0.989, 1.192)
Observations	185	185	185	185	185	185

Note: *p < .05; **p < .01; ***p < .001.

TABLE 10 Logistic regression models for proximal violent acts, 7-point scoring.

Main effects OR (95% CI)						Fully adjusted AOR (CI)
C1	0.987					0.813
	(0.790, 1.183)					(0.569, 1.057)
C2		1.292***				1.152
		(1.140, 1.444)				(0.958, 1.346)
C3			1.006			0.929
			(0.869, 1.142)			(0.771, 1.087)
C4				1.428***		1.221
				(1.239, 1.617)		(0.967, 1.476)
C5					1.371***	1.288***
					(1.181, 1.561)	(1.045, 1.532)
Constant	0.458	0.207	0.428*	0.107	0.152	0.154
	(−0.361, 1.277)	(−0.353, 0.768)	(−0.036, 0.892)	(−0.740, 0.955)	(−0.579, 0.883)	(−0.923, 1.231)
Observations	208	208	208	208	208	208

Note: *$p < .05$; **$p < .01$; ***$p < .001$.

TABLE 11 Logistic regression models for proximal violent acts, 3-point scoring.

Main effects OR (95% CI)					Fully adjusted OR (CI)
R1	0.912				0.618
	(0.501, 1.324)				(0.137, 1.100)
R2		1.239			1.188
		(0.828, 1.650)			(0.710, 1.667)
R3			1.587***		1.473
			(1.066, 2.107)		(0.874, 2.072)
R4				2.194***	2.430***
				(1.655, 2.733)	(1.803, 3.058)
R5					1.252
					(0.656, 1.848)
					0.844
					(0.128, 1.561)
Constant	0.465*	0.333	0.218	0.145	0.293
	(−0.059, 0.988)	(−0.234, 0.900)	(−0.615, 1.051)	(−0.690, 0.980)	(−0.743, 1.329)
					0.129
					(−1.090, 1.349)
Observations	185	185	185	185	185

Note: *$p < .05$; **$p < .01$; ***$p < .001$.

TABLE 12   Logistic regression models for proximal violent acts, 7-point scoring.

Main effects OR (95% CI)				Fully adjusted OR (CI)
R1	1.059			0.945
	(0.895, 1.222)			(0.750, 1.140)
R2		1.171***		1.166
		(1.015, 1.327)		(0.974, 1.359)
R3			1.132	1.052
			(0.929, 1.334)	(0.819, 1.285)
R4			1.390***	1.452***
			(1.168, 1.612)	(1.183, 1.722)
R5			1.050	0.790
			(0.827, 1.274)	(0.498, 1.081)
Constant	0.367	0.266	0.263 / 0.123	0.349 / 0.178
	(−0.208, 0.941)	(−0.319, 0.851)	(−0.622, 1.148) / (−0.812, 1.058)	(−0.696, 1.393) / (−1.058, 1.413)
Observations	208	208	208 / 208	208 / 208

Note: *p < .05; **p < .01; ***p < .001.

Increasing the number of responses available appeared to improve the consistency of item scores across Historical, Clinical and Risk domains, and particularly so for Historical items. At best, this supported the notion that increasing the number of scoring options, diversifying anchor point descriptions and broadening the definition of Presence items allowed greater freedom for the clinician to rate violence potential in item scores. This adds to the literature that an increased scale length improves scale validity (Chang, 1994; Hancock & Klockars, 1991; Loken et al., 1987) and better reflects the clinician's opinion of the factor they are rating (Preston & Colman, 2000).

The composite Historical domain reflected the greatest difference between the 3- and 7-point formats. Where it is easier to score highly on an item (e.g. history of violence), the longer 7-point response format was able to provide more information about risk factors relative to the standard 3-point format. In practice, this meant clinicians could identify aspects of violence potential which the shorter format did not allow. For example, a rater may have scored highly on previous history of violence on both the 3- and 7-point formats, but the longer format allowed the clinician to reflect differences in the extent to which previous violence was associated with future violence, in respect of frequency, chronicity and severity, compared to the shorter format. This could enable finer grain evaluations of risk factors, which have utility in distinguishing risk between service users and increasing the sensitivity of the measure to detect change over time for individual cases. If the 7-point format yielded more useful information about violence potential, then using a longer response format has the potential to further the knowledge and understanding of violence risk.

The criterion-related, or predictive, validity of the 7-point response format was comparable with the 3-point version, with the former marginally stronger at predicting violence and negative outcomes overall. The greatest difference was seen in the predictive validity of historical items, with the 7-point format recording four significant predictors and the 3-point model only one. In the adjusted model, only the 7-point model retained a single significant predictor. For the clinical and risk items, the 3- and 7-point formats performed similarly. These results may reflect the complexity of measuring violence potential, with prediction improved using a longer response format, allowing for better detection of subtle differences and complexity, particularly in historical items. The predictive validity for both response formats was similar to those reported in meta-analyses of violence risk measures (e.g. Fazel et al., 2012). It should be noted that some negative outcomes, such as loss of enhanced security status, were proxy indicators of violence and not directly observed instances. This may have over- or underestimated prevalence of violence in the sample. However, the definitions of negative outcomes were the same for both response formats and should not have affected differences observed between the two.

Study strengths

This study used a combination of validity tests, including a robust analysis of predictive validity using logistic regression. The results of analyses for 3- and 7-point response formats generally concurred, lending credibility to the positive findings in relation to the performance of the 7-point format. The study incorporated post-hoc analyses to control for the increased number of tests introduced with multiple points on a longer response format, reducing the probability of false positive results. Significant findings remained after these tests, increasing confidence in the reliability of the results. This strengthened the interpretation that the 7-point format performed at least as well, if not marginally better in some areas, than its 3-point counterpart.

Limitations

This study considered only the rating of Presence items using both 3- and 7-point formats. Therefore, some caution should be adopted when extrapolating these analyses to relevance ratings. Due to the low

frequency of distal violence, the predictive validity analysis makes use of violent behaviour indicators that may have taken place anytime up to 6 months from the recorded rating. This is potentially less reliable than limiting outcomes to observed violence occurring in closer proximity to the HCR20 rating.

Constraints on generality

This study did not collect identifiable data on the service participants in prisons and secure health units, and it is therefore not possible to comment directly on the representation of diversity factors in the sample. General demographics of the participating sites were included, and those selected for the study comprised a relatively large sample of this group. It is well documented that white service users are over-represented in psychological interventions in offender settings. The study therefore may not generalize to the wider prison and secure health population.

Implications for further research and practice

The 7-point response format had similar, if not better, predictive validity, internal consistency, and interrater reliability compared to the standard 3-point format. At best, it may enhance the properties of the HCR-20 V3 by enabling clinicians to make more nuanced judgements about violence risk. This could have clinical utility in improving the detection of changes in risk factors over time at the individual level. This has positive ramifications for making judgements about improvement or deterioration in risk, impacting on clinical and legal decision-making about service progression or enhanced restrictions; and as a consequence, their more appropriate use in public protection.

Further research would be needed to determine whether detecting incremental changes in risk over time would be reflected in significant changes in risk scores at both the individual and group levels. If a longer response format enabled detection of significant change in risk scores for services, it could have important implications for determining the suitability of treatment and containment models used by services. This could make an important contribution to the evaluation of service effectiveness and enable more informed decisions about ongoing programme funding.

Finally, if future research suggested the novel 7-point format was superior to the standard 3-point version, it could make a significant contribution to the evidence base and our understanding of the nuances and complexities of structured professional judgement risk assessment. If the 7-point response format of the HCR20 V3 represented significant improvements over its predecessor, it would indicate the measure better reflects violence potential.

Conclusion

This study aimed to test the validity of a longer 7-point response format on the HCR-20 V3, by comparing its performance to the existing 3-point version. The findings suggested the longer format performed as well as, and no worse than, the original version. This supports its potential utility for measuring violence risk, which was beyond the scope of this study. There was some indication the longer response format allowed clinicians to record more nuanced aspects of violence risk, which the shorter 3-point format does not permit. A greater capacity to record gradual increments in violence potential may improve clinicians' ability to detect changes in risk over time. This could enhance service user and professional optimism about the potential for change. Increasing the capacity to detect significant change at the individual and group level has important implications for clinical and legal decision-making, and could help determine programme and service effectiveness. Further research is necessary to establish whether the 7-point response format is more likely than its 3-point counterpart to detect changes in risk over time.

AUTHOR CONTRIBUTIONS

Philip Minoudis: Conceptualization; methodology; supervision; writing – original draft; writing – review and editing. **Bardana Singh:** Data curation; investigation. **Chrissie Lane:** Investigation; data curation. **Marvin Iroegbu:** Investigation; data curation. **Caitriona Higgins:** Investigation. **Clare Wellington:** Investigation. **Stephen D. Hart:** Supervision; writing – review and editing. **Kevin S. Douglas:** Conceptualization; supervision. **Mark Freestone:** Formal analysis; writing – review and editing.

FUNDING INFORMATION

CONFLICT OF INTEREST STATEMENT

Stephen D. Hart and Kevin S. Douglas declare that they are authors and receive royalties from sales of the HCR-20 V3. The remaining authors do not declare any conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Philip Minoudis https://orcid.org/0009-0001-5425-2022
Bardana Singh https://orcid.org/0000-0002-2695-5601
Marvin Iroegbu https://orcid.org/0009-0001-8091-0738
Stephen D. Hart https://orcid.org/0000-0002-1027-6081
Kevin S. Douglas https://orcid.org/0000-0002-4486-9918
Mark Freestone https://orcid.org/0000-0002-0992-8468

REFERENCES

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, *20*(1), 40–49.

Belfrage, H., & Douglas, K. S. (2002). Treatment effects on forensic psychiatric patients measured with the HCR-20 violence risk assessment scheme. *International Journal of Forensic Mental Health*, *1*(1), 25–36. https://doi.org/10.1080/14999013.2002.10471158

Buuren, S. V., & Groothuis-Oudshoorn, K. (2010). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–68.

Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, *18*(3), 205–215. https://doi.org/10.1177/014662169401800302

Cox, E. P., III. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research*, *17*(4), 407–422.

Colman, A. M., Norris, C. E., & Preston, C. C. (1997). Comparing rating scales of different lengths: Equivalence of scores from 5-point and 7-point scales. *Psychological Reports*, *80*(2), 355–362. https://doi.org/10.2466/pr0.1997.80.2.355

De Vries Robbé, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. I. (2015). Changes in dynamic risk and protective factors for violence during inpatient forensic psychiatric treatment: Predicting reductions in postdischarge community recidivism. *Law and Human Behavior*, *39*(1), 53–61. https://doi.org/10.1037/lhb0000089

Department of Health UK. (2007). *Best practice in managing risk: Principles and evidence for best practice in the assessment and Management of Risk to self and others in mental health services*. Department of Health.

Douglas, K. S., & Shaffer, C. S. (2020). The science of and practice with the HCR-20 V3 (historical-clinical-risk management-20, version 3). In *Handbook of violence risk assessment* (pp. 253–293). Routledge.

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20V3: Assessing risk of violence – User guide*. Mental Health, Law and Policy Institute, Simon Fraser University.

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: systematic review and meta-analysis. *BMJ*, *345*, e4692.

Fleiss, J. L., Levin, B., & Paik, M. C. (1981). The measurement of interrater agreement. *Statistical Methods for Rates and Proportions*, *2*(212-236), 22–23.

Ghiselli, E. E. (1955). A scale for the measurement of initiative. *Personnel Psychology*, *8*(2), 157–164.

Green, P. E., & Rao, V. R. (1970). Rating scales and information recovery - how many scales and response categories to use. *Journal of Marketing*, *34*(3), 33–39. https://doi.org/10.2307/1249817

Hancock, G. R., & Klockars, A. J. (1991). The effect of scale manipulations on validity: Targetting frequency rating scales for anticipated performance levels. *Applied Ergonomics*, *22*(3), 147–154.

Hart, S. D., & Douglas, K. S. (2019). *Violence risk assessment and management: From prediction to prevention. Risk rules: A practical guide to structured professional judgment and violence prevention* (pp. 87–92). Pavilion Publishing and Media.

Hart, S. D., & Douglas, K. S. (2023). *Violence risk assessment and management* (pp. 307–325). The Oxford Handbook of Psychology and Law.

Hart, S. D., Douglas, K. S., & Guy, L. S. (2016). The structured professional judgement approach to violence risk assessment: Origins, nature, and advances. In D. P. Boer, A. R. Beech, T. Ward, L. A. Craig, M. Rettenberger, L. E. Marshall, & W. L. Marshall (Eds.), *The Wiley handbook on the theories, assessment and treatment of sexual offending* (Vol. *2*, pp. 643–666). Wiley Blackwell.

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.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Loken, B., Pirie, P., Virnig, K. A., Hinkle, R. L., & Salmon, C. T. (1987). The use of 0-10 scales in telephone surveys. *Journal of the Market Research Society*, *29*(3), 353–362.

Michel, S. F., Riaz, M., Webster, C., Hart, S. D., Levander, S., Müller-Isberner, R., Michel, S. F., Hart, S. D., Tiihonen, J., Repo-Tiihonen, E., Tuninger, E., & 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. https://doi.org/10.1080/14999013.2012.760182

Monahan, J., Steadman, H. J., Silver, E., Applebaum, P., Robbins, P., Mulvey, E., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. Oxford Univeristy Press.

Pasta, D. J. (2009). Learning when to be discrete: Continuous vs. categorical predictors. *SAS Global Forum 2009 Proceedings*, Paper 248–2009. http://support.sas.com/resources/papers/proceedings09/248-2009.pdf

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15.

R Core Team. (2013). R: A language and environment for statistical computing.

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., Ho, R. M. Y., Large, M. M., Nielsen, L. H., Pham, T. H., Rebocho, M. F., Reeves, K. A., Rettenberger, M., de Ruiter, C., Seewald, 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. https://doi.org/10.1080/14999013.2014.922141

Symonds, P. M. (1924). On the loss of reliability in ratings due to coarseness of the scale. *Journal of Experimental Psychology*, *7*(6), 456.

Verschueren, S., Jeandarme, I., Libijn, I., & Bogaerts, S. (2023). Longitudinal change of risk factors during treatment in high-security forensic psychiatry in Flanders. *International Journal of Offender Therapy and Comparative Criminology*, *67*, 1722–1740. https://doi.org/10.1177/0306624X221144293

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk of violence* (2nd ed.). Mental Health Law & Policy Institute, Simon Fraser University.

Webster, C. D., Eaves, D., Douglas, K. S., & Wintrup, A. (1995). *The HCR-20 scheme: the assessment of dangerousness and risk–version 1*. Simon Fraser University and Forensic Psychiatric Services Commission of British Columbia.

Wilson, C. M., Desmarais, S. L., Nicholls, T. L., Hart, S. D., & Brink, J. (2013). Predictive validity of dynamic factors: Assessing violence risk in forensic psychiatric inpatients. *Law and Human Behavior*, *37*(6), 377–388. https://doi.org/10.1037/lhb0000025

Wilson, K., Freestone, M., Taylor, C., Blazey, F., & Hardman, F. (2014). Effectiveness of modified therapeutic community treatment within a medium-secure service for personality-disordered offenders. *The Journal of Forensic Psychiatry & Psychology*, *25*(3), 243–261. https://doi.org/10.1080/14789949.2014.908317

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1.