

## RESEARCH ARTICLE

WILEY

# Evaluating the HCR-20<sup>V3</sup> violence risk assessment measure with mentally disordered offenders and civil psychiatric patients in China

Yifan Chen<sup>1</sup> | Kevin S. Douglas<sup>2,3,4</sup> | Zhuo Zhang<sup>5</sup> | Cunli Xiao<sup>6</sup> | Haiyan Wang<sup>7</sup> | Yuhao Wang<sup>5</sup> | Ai Ma<sup>5</sup>

<sup>1</sup>School of Criminal Investigation, People's Public Security University of China, Beijing, China

<sup>2</sup>Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada

<sup>3</sup>Centre for Research and Education in Forensic Psychiatry, Helse Bergen-HF, Bergen, Norway

<sup>4</sup>Centre for Research and Education in Forensic Psychiatry, Oslo University Hospital-HF, Oslo, Norway

<sup>5</sup>School of Sociology, China University of Political Science and Law, Beijing, China

<sup>6</sup>Beijing Xicheng Ping'an Hospital, Beijing, China

<sup>7</sup>Department of Experimental and Applied Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

## Correspondence

Ai Ma, School of Sociology, China University of Political Science and Law, 25 Xitucheng Road, Haidian District, Beijing 100088, China. Email: [chinafzxl@163.com](mailto:chinafzxl@163.com)

## Funding information

Fundamental Research Funds for the Central Universities, Grant/Award Numbers: 2022JKF427, 20ZFD82002

## Abstract

The current prospective risk assessment study evaluated the application of the Chinese translation of the Historical-Clinical-Risk Management-20 Version 3 (HCR-20<sup>V3</sup>) in a sample of 152 offenders with mental disorders and civil psychiatric patients. The ratings of the presence and relevance of risk factors were compared, as well as summary risk ratings (SRRs), both across offenders and civil psychiatric patients, and across male and female sub-samples. Interrater reliability was consistently "excellent" for the presence and relevance of risk factors and for SRRs. Concurrent validity analyses indicated that HCR-20<sup>V3</sup> was strongly correlated with Violence Risk Scale (from  $r = 0.53$  to  $0.71$ ). The results of predictive validity analyses provided strong support for the bivariate associations between the main indices of HCR-20<sup>V3</sup> and violence within 6 weeks, 7–24 weeks, and 6 months; SRRs added incrementally to both relevance and presence ratings across three follow-up lengths.

## KEYWORDS

civil psychiatric, HCR-20<sup>V3</sup>, offender, risk assessment, violence

## 1 | INTRODUCTION

In recent years, violent incidents caused by people with mental disorders have occurred frequently in China. According to a national survey conducted by the Ministry of Public Security of China in 2014, 2.1 million people with severe mental disorders were registered in China. Of these people, 0.11 million were responsible for various types

of violence. At a formal meeting jointly held by the Ministry of Civil Affairs, the National Health Commission, and the China Disabled Persons' Federation in 2018, it was announced that the number of patients with severe mental disorders registered in China had reached 5.81 million. A latest nationwide epidemiological survey of mental disorders in China founded that the prevalence of schizophrenia was 0.6% (Huang et al., 2019). According to the population data from China's National Bureau of Statistics, the total population of mainland China is now 1.41 billion. For schizophrenia alone, patients were estimated to exceed 8.48 million in China. It can be seen that the number of patients with serious mental disorders is increasing year by year at a high rate. The risk assessment of violence for such an enormous number of patients with mental disorders has become a very urgent and vitally important social issue in China.

## 1.1 | Relationship between mental disorders and violence

The general public perceives people with mental disorders as violent and highly dangerous. According to a study on the attitude of the general public in China toward people with mental disorders, more than 60% of the general population and more than 55% of mental hospital personnel believe that people with mental disorders are violent (Gao & Fei, 2001). Indeed, such beliefs are common throughout the world (Douglas, 2019; Monahan, 1992).

Although there remains some controversy and disagreement about the importance of mental disorder as a risk factor for violence (Bonta et al., 2014; Guy & Douglas, 2015), a meta-analysis study indicates an increased occurrence of violence among people with major mental disorder compared to those without (Douglas et al., 2009), despite it likely being the case that most people with major mental disorders do not engage in violence. Regardless of the precise nature of the association, it is imperative to evaluate risk for violence among people with major mental disorders in certain contexts. For some people, mental disorder is essential for understanding their risk, whereas for others, it is not. This comes to the point of risk assessment—to determine when certain risk factors are, or are not, relevant to a person's violence (Douglas et al., 2013).

One point of relative agreement is that risk factors for violence are highly comparable for people with and without mental disorder (Bonta et al., 2014; Heilbrun et al., 1998). As such, it is imperative to consider common risk factors, such as anti-social attitudes, a lack of education, impulsivity, reasoning and problem-solving abilities (Andrews et al., 2006; Blackburn, 2004), in addition to features of mental disorder (Douglas et al., 2009, 2013), in order to complete a comprehensive violence risk assessment. This approach helps to not only protect the personal safety of the public but also to erase the stigma that the general public attached to mentally ill persons so that the personal rights and interests of people with mental disorders who pose no threat of violence are guaranteed.

## 1.2 | Risk assessment tools and HCR-20

As a risk assessment tool developed for people with or without mental disorders, the Historical, Clinical, Risk Management-20 (HCR-20) is one of the most widely used. Its first edition was developed in 1995 (Webster et al., 1995), and 2 years later, the second edition was introduced (Webster et al., 1997). Over the following decades, the HCR-20 has been translated into many languages, including more than 20 languages officially authorized by the developers, and it has been widely disseminated, applied, and accepted across over 40 countries worldwide. For example, Singh et al. (2014) surveyed 2135 clinicians across 44 countries spanning Europe, North America, Oceania, Asia, South America, and Africa and found that the HCR-20 was the most commonly used violence risk assessment tool (among the top 12 ranked assessment tools). Of the respondents, 1032 (48.34%) had used the HCR-20, and 669 (31.33%) had used it over the past year.

The authors revised the HCR-20 since the second edition was released, and the third and latest edition (HCR-20<sup>V3</sup>) was released in 2013 (Douglas et al., 2013). Numerous studies conducted across different national, cultural, and ethnic conditions have confirmed that the HCR-20<sup>V3</sup> has a robust predictive power with regard to

violence among people with or without mental disorders with satisfactory reliability and validity (e.g., Adjorlolo & Chan, 2019; Blanchard & Douglas, 2011; de Vries Robbé & de Vogel, 2010; Douglas & Belfrage, 2014; Doyle et al., 2013, 2014; Eidhammer et al., 2013; Green et al., 2016; Hogan & Olver, 2016, 2019; Howe et al., 2016; Kötter et al., 2014; Nijdam-Jones et al., 2021, 2022; Penney et al., 2016; Persson et al., 2017; Sea & Bang, 2021; Smith & Edens, 2013; Strub et al., 2014). To date, most of the studies with the HCR-20<sup>V3</sup>—like with the HCR-20 Version 2—have been conducted in Western countries (de Vogel et al., 2022). This trend is gradually changing with research from Non-Western countries, like South Korea (Sea & Bang, 2021), Ghana (Adjorlolo & Chan, 2019), and Mexico (Nijdam-Jones et al., 2021, 2022). Thus far, however, research on the reliability and validity of the authorized Chinese translation of the HCR-20<sup>V3</sup> has not been reported.

Furthermore, the performance of violence risk assessment measures, in general, and the HCR-20<sup>V3</sup>, specifically, across gender, remains unclear. Some previous studies have indicated that the rate of violence among mentally ill male patients differs from that among female patients, and the relative ratio of these two rates differs across different follow-up lengths. For instance, a 6-month follow-up study showed that the rate of violence among male patients with mental disorders was higher than that among female patients, but the situation was reversed over a 12-month follow-up period (Doyle et al., 2014). Persson et al. (2017) also conducted a 12-month follow-up study and found that the rate of violence among female patients with mental disorders was higher than that among male patients; as the follow-up time increased to 15.5 months, the rate of violence among female patients with mental disorders remained higher than that among male patients (Green et al., 2016). Therefore, investigations of the interaction between follow-up length and gender with regard to the predictive power of violence risk assessment tool are of great significance. In addition, the rate of violence among patients at forensic psychiatric hospitals may differ from that among imprisoned offenders depending on the follow-up length (Persson et al., 2017).

Strub et al. (2014) found that the gender and sub-sample type (i.e., offenders vs. civil psychiatric patients) of people with mental disorders did not affect the predictive power of each major index of the HCR-20<sup>V3</sup> over different assessment periods. Up to now, there is no empirical study on the use of HCR-20<sup>V3</sup> in Chinese population, whether this conclusion remains valid for Chinese patients with mental disorders requires empirical confirmation.

### 1.3 | Current study

This is the first study of the official Chinese translation of the HCR-20<sup>V3</sup>, and indeed the first study of *any* approved Chinese translation of the HCR-20. Past studies of *unauthorized* translations of past versions of the HCR-20 have been summarized (Gu et al., 2014; Zhou et al., 2016), although the original studies, and the reviews based on them, ought not to be given credibility or weight because the translations were not approved, or even known about, by the HCR-20 authors. There are some obvious defects in unauthorized translations of the HCR-20 and previous studies using these tools: (1) inappropriate translations may result in misunderstandings of specific items, (2) assessments without formal training may result in low reliability and validity, and (3) negative impacts on academic ethics. Here, our goal is to provide an authorized Chinese translation of the HCR-20<sup>V3</sup>, in which the items are translated appropriately and the assessment implementers are professionally trained.

The current study introduces the official Chinese translation of the HCR-20<sup>V3</sup>, localizes it through empirical study, and examines its reliabilities and validities with regard to the social and cultural environment of China. Furthermore, its actual predictive power regarding the risk of violence among people with mental disorders is examined.

Specifically, we created the Chinese translation of the HCR-20<sup>V3</sup> through a back-translation procedure. We used the HCR-20<sup>V3</sup> and the Violence Risk Scale (VRS) to assess the violence risks of patients with mental disorders from hospitals and prisons across gender and two sub-samples (offenders with mental disorders and civil psychiatric patients). More specifically, we tested the interrater reliability, concurrent, and predictive validity of the HCR-20<sup>V3</sup>.

In addition, in reference to the standards of Strub et al. (2014), we designated 6 weeks after the baseline interview as the “short term” and 6 months after the baseline interview as the “long term.” To further investigate the predictive

power of the HCR-20<sup>V3</sup> with regard to violence among people with mental disorders over different follow-up lengths, we divided the long-term into two phases. The phase of the first 6 weeks after the baseline interview (i.e., “short term”) was regarded as the “near phase,” and the phase from the seventh week to the 24th week was regarded as the “far phase.” Over these three follow-up lengths (i.e. 6 weeks, 7–24 weeks, and 6 months), we examined the predictive validity of each major index of the HCR-20<sup>V3</sup> as well as tested whether the gender and sub-sample of people with mental disorders affect the predictive accuracy of the tool. Moreover, we compared the predictive power of major indices to determine whether they differed.

## 2 | METHOD

### 2.1 | Participants

This study recruited individuals from five wards of two psychiatric hospitals and two monitoring areas of one prison especially for offenders with mental illness in Beijing, based on the following inclusion criteria: (1) age over 18 years but below 75 years; (2) diagnosis of a mental disorder by a psychiatric hospital or forensic psychiatric assessment agency based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), with or without comorbidities; and (3) signed informed consent. The exclusion criteria were as follows: (1) incapable of verbal communication due to impaired intelligence or disability, (2) currently experiencing highly unstable psychiatric symptoms with serious hostility, and (3) unable to comply with the assessment.

Based on the above criteria, 152 patients with mental disorders (94 males and 58 females) with an average age of 48.75 years (range 23–75) were recruited for this study; of these patients, 93 were from hospitals and 59 were from prisons. The average age of the male and female patients was 49.20 years (range 23–74) and 48.02 years (range 25–75), respectively. The average age of patients from hospitals and prisons was 49.68 years (range 25–75) and 47.29 years (range 23–74), respectively. Statistical tests revealed that the average ages of male and female participants and the ages of participants from hospitals and prisons were not significantly different.

Using the DSM-5, the four most common types of mental disorder among the current participants were schizophrenia spectrum and other psychotic disorders ( $n = 109$ , 71.7%), depressive disorders ( $n = 26$ , 17.1%), personality disorders ( $n = 16$ , 10.5%), and bipolar and related disorders ( $n = 15$ , 9.9%). The full description of diagnostic composition of participants appears in Table S1 in the Supporting Information available online.

The study was approved by the Prison Administration Bureau of Beijing and the Institutional Research Ethics Committee of Beijing Xicheng Ping'an Hospital.

### 2.2 | Measures

#### 2.2.1 | HCR-20<sup>V3</sup>

The HCR-20<sup>V3</sup> retains the three subscales of the HCR-20<sup>V2</sup> and its 20-item structure. However, various modifications were made to specific items. For example, the result of the Psychopathy Checklist-Revised (PCL-R) was no longer required (although it may still be used), and several sub-items were added to multiple items. Moreover, the HCR-20<sup>V3</sup> followed the Guideline for Stalking Assessment and Management (SAM; Kropp et al., 2008) and added indicators to the definition of each item and sub-item (i.e., several examples were provided to explain the situation or case that a given risk factor measure) to help improve the interrater reliability of the tool.

Unlike some actuarial risk assessment instruments, the HCR-20<sup>V3</sup> does not identify a score cutoff to classify assessed persons with different levels of risk; instead, evaluators generate an overall assessment of the participant's level of risk after completing the first six steps and completing the summary risk rating (SRR) regarding the participant

across three levels of risk: low, moderate, and high. In addition, the HCR-20<sup>V3</sup> allows evaluators to perform SRRs on the participant with regard to serious physical harm and imminent violence.

In this study, we used the Chinese translation of the HCR-20<sup>V3</sup> to assess the risk of violence for Chinese patients with mental disorders. Since HCR-20<sup>V3</sup> is a well-established instrument, we strictly followed the original structure and items of HCR-20<sup>V3</sup>. To avoid cultural differences between Western and East Asian culture, we examined whether the original items were inconsistent with Chinese society. To ensure the effectiveness of assessment process, we followed the seven-step procedure suggested by the guidelines of the HCR-20<sup>V3</sup>. In Step 2 (i.e., determine presence of risk factors), we needed to decide whether the item definition has been met, or satisfied, with regard to the participants. In Step 3 (i.e., determine relevance of risk factors), we assessed to what extent each of the items is relevant to the past occurrence or future likelihood of violence with regard to the participant (i.e., has played a vital role in causing past violence or is likely to do so in any future violence). In Step 7, we also completed the SRR. Since the present study focuses on general violence, neither serious physical harm nor imminent violence was evaluated in this study.

### 2.2.2 | VRS

The VRS is developed by Wong and Gordon (1999-2003) to assess risk for violence and to help guide treatment plans. Based on the risk-need-responsivity (R-N-R) principle and criminal behavioral psychology (Andrews & Bonta, 1994), the VRS is composed of six static factors and 20 dynamic factors. Each item is scored with a 4-point scale in which "0" represents "there is no problem regarding this item," and "3" represents "there is a serious problem regarding this item." In this study, the Chinese translation of Violence Risk Scale (VRS-C; Zhang et al., 2012) was used in tests of concurrent validity.

### 2.2.3 | Outcome measures

In this study, violence was defined as "actual, attempted, or threatened infliction of bodily harm to another person," where "bodily harm includes both physical and serious psychological harm" and "psychological harm includes fear of physical injury, and other emotional, mental or cognitive consequences of the act in question" (Douglas et al., 2013, p. 36). Violence was assessed according to the nine types of violence and other aggressive acts used in the MacArthur Study of Mental Disorder and Violence, including throwing objects, push-grab-shove, slap, kick-bite-choke, hit-beat up, force sex, weapon threat, weapon use, and other unknown type (Monahan et al., 2001). Three sources of information were used to ascertain the occurrence and details of a violent incident: (1) interviews with participants; (2) interviews with collateral individuals (i.e., someone who would know what was going on in participants' lives, such as participants' attending physicians, nurses, social workers, support staff, police officers, family members, and other patients or prisoners within the institution); (3) reviews of all available records at the institution (medical records and work records of the prison staff). These three sources of information were all coded and compared. Two face-to-face interviews (6 weeks and 6 months later) and at most six telephone interviews (once a month) were conducted after the baseline interview. The participants and collateral individuals were asked whether the participant had engaged in several types of violent acts during each follow-up period. If the interviewer got positive responses, then more details regarding this incident were asked, including: (1) the seriousness of the violence, (2) who was involved, (3) when and where the incident occurred, and (4) why this incident happened.

## 2.3 | Procedure

### 2.3.1 | Translation

After receiving authorization from the authors of the HCR-20<sup>V3</sup>, three graduate students in forensic psychology independently translated the HCR-20<sup>V3</sup> into Chinese. The draft of the Chinese translation of the HCR-20<sup>V3</sup> was

obtained by comprehensively comparing the three translations via a joint review performed by two psychology professors. The inaccurate translations were revised and improved. A bilingual expert fluent in both Chinese and English back-translated the revised version into English. The above process was repeated several times to ensure that the Chinese translation accurately reflected the original meaning of the English version while conforming to Chinese expression styles, ultimately creating the Chinese translation of the HCR-20<sup>V3</sup>.

There were some challenges in translating the HCR-20<sup>V3</sup> from English to Chinese. For example, the fifth risk factor of the HCR-20<sup>V3</sup> is the "History of Problems with Substance Use" (H5). This risk factor pertains to a history of problems with the use of, or dependence on mind- or mood-altering substances, including alcohol, illicit drugs, licit drugs, or incidental substances. However, there is no word that can adequately describe "Substance Use" in Chinese. Besides involvement in drugs trade is one of the most serious crimes in China, taking drugs (including marijuana) is also an act that can face severe punishment. In this study, no subjects participated in drug use and drug trade. Therefore, we mainly interpreted the substance use as the abuse of alcohol. Another example is the risk factor "Future Problems with Professional Services and Plans" (R1). In the HCR-20<sup>V3</sup>, professional services are defined as those health care, social service, educational, vocational, and criminal justice programs designed to manage violence risk. Because violence risk assessment is a relatively new research field in China, such professional services are still underdeveloped. This also brought some challenges to our translation. The translation process and the final product were approved by the lead author of the HCR-20<sup>V3</sup>.

### 2.3.2 | Research assistant (RA) training

Four graduate students in forensic psychology served as RAs to systematically study the manual of the HCR-20<sup>V3</sup>. The RAs also completed a 20-h online HCR-20<sup>V3</sup> training course. Moreover, they received face-to-face training from psychiatrists to acquire mental illness-related knowledge, understand how to interview patients with mental disorders, and take the necessary precautions. They also received an intensive training session concerning the use of the VRS from trainers authorized by the author of the VRS, in which they studied how to assess the VRS items and their scoring rules. After completing these trainings, four RAs administered the HCR-20<sup>V3</sup> and VRS to four patients with mental disorders and independently scored the assessment items. Assessment results that differed among the RAs were discussed based on the unified evaluation criteria that were ultimately established.

### 2.3.3 | Recruitment and baseline interviews

The assessment procedures were as follows. (1) The participant was verbally informed of the purpose of the research program, and his or her informed consent was obtained by signing the relevant documents. (2) Two RAs jointly reviewed the clinical case or forensic psychiatric decision concerning the participant; they also took notes on the status of the mental disorder as well as the participant's past histories of violence and hospitalization. (3) The RAs briefly interviewed the psychiatrists, ward nurses, or prison ward officials of the institution to which the participant belonged to understand whether the participant was still showing mental disorder symptoms and how the participant behaved within the institution. In a small number of cases, family members of the participant were also briefly contacted to clarify missing or inconsistent information. (4) In a separate and comfortable environment, a semi-structured interview was conducted with the participant based on the interview outline. After completing the interview, some daily necessities were presented to the participant to express our gratitude for his or her contribution to this study. (5) Two RAs independently scored each of the HCR-20<sup>V3</sup> and VRS items with regard to the participant based on their interview notes and the information collected. (6) A score table was then collated, and data entry was considered as complete. In addition, the interrater reliability was analyzed based on the assessment results of the three RAs across 33 participants using the HCR-20<sup>V3</sup>. Except for conducting independent interviews and taking separate notes, the three RAs followed the same procedures as those described above regarding the baseline interview.

TABLE 1 Description of participants and retention.

Phases	Participants
Baselines	152
Male	94
Female	58
Civil psychiatric	93
Offender	59
First follow-up visit	151
Male	94
Female	57
Civil psychiatric	92
Offender	59
Second follow-up visit	149
Male	92
Female	57
Civil psychiatric	92
Offender	57

### 2.3.4 | Reassessment and missing follow-up assessments

Two follow-up visits were performed for each participant 6 weeks and 6 months after the baseline interview. In addition, each participant was followed up with each month by RAs over the phone. The above follow-ups were conducted to determine whether the participant had engaged in violence during the follow-up period.

The procedure of the follow-up visit was as follows. First, all available records at the institution were reviewed. The review process focused on various pieces of information concerning the patient's violence and aggression as well as situations such as whether the patient had been restrained or punished (e.g., confined, denied the opportunity to smoke, restricted from traveling, and so on). Because these situations often indicate unruly patient behavior, their causes must be examined with the help of medical or prison staff to screen for acts of violence that had been committed but not recorded.

Second, if the participant was still at one of the institutions, then he or she would be interviewed. First, RAs encouraged the participant to report whether he or she had committed any violent acts since the previous interview and whether he or she had been subjected to any violence or aggression from other patients within the institution. All of the information reported by the participant required double checking. For example, staff members would be interviewed to determine whether they had witnessed or heard about the incidents reported by the participant; medical records or work records would be checked; and all of the perpetrators, victims, and bystanders would be interviewed about the occurrence and details of a violent incident.

In the event that the participant had already left the institution (i.e., discharged from the hospital or released from prison), the participant was interviewed over the phone. First, the time when the participant left the institution was determined by looking up his or her file, and the participant's history of violence before leaving the institution was examined in accordance with the above procedure. The participant's history of violence after leaving the institution was examined via telephone interviews with the participant and his or her family members or guardians. Participants who were impossible to contact after leaving their institutions were regarded as "lost."

Finally, all of the above sources of information were compared and integrated, and the cases of violence of each patient within 6 weeks ( $n = 151$ ), 7–24 weeks ( $n = 149$ ), and 6 months ( $n = 149$ ) after baseline were coded. Table 1

describes participants and retention on different phases. The presence of one or more violent acts within the period was coded as “1”; no violent acts were coded as “0.”

### 3 | RESULTS

#### 3.1 | Descriptive statistics

The HCR-20<sup>V3</sup> item ratings for participants are shown in Table S2. The means and standard deviations of patients on the HCR-20<sup>V3</sup> and VRS indices are also shown in the Supporting Information (for details, see Table S3). With regard to the SRR assessment result, 18 (11.8%) participants were assessed as “*high risk*,” 40 (26.3%) as “*moderate risk*,” and 94 (61.8%) as “*low risk*.”

#### 3.2 | Rate of violence

The rate of violence of all participants in the near phase (i.e., over the short term) was 16.6% (25 of 151), that of those in the far phase was 22.1% (33 of 149), and that of those 6 months after baseline interview (i.e., over the long term) was 28.9% (43 of 149). The changes in this rate are shown in Figure S1. Figure S1 shows that the rate of violence increased as the follow-up time increased.

During the near phase, among the 25 participants who had one or more violent incidents, 15 (15 of 94, 16.0%) were men, and 10 (10 of 57, 17.5%) were women; 13 (13 of 92, 14.1%) were from hospitals, and 12 (12 of 59, 20.3%) were from prisons. During the far phase, among the 33 participants who had one or more violent incidents, 23 (23 of 92, 25%) were men, and 10 (10 of 57, 17.5%) were women; 18 (18 of 92, 20.0%) were from hospitals, and 15 (15 of 57, 26.3%) were from prisons. And during the entire 6 months after the baseline interview, among the 43 participants who had one or more violent incidents, 28 (28 of 92, 30.4%) were men, and 15 (15 of 57, 26.3%) were women; 24 (24 of 92, 26.1%) were from hospitals, and 19 (19 of 57, 33.3%) were from prisons. The changes in rates of violence among two genders and settings across different time lengths are shown in Figures S2 and S3. Figure S2 shows that the rate of violence among male patients with mental disorders was lower than that among female patients during the near phase, but the situation was reversed both in 7–24 weeks and 6 months follow-up periods. Figure S3 displays that the rate of violence among civil psychiatric patients was always lower than that among offenders in the three follow-up lengths, and the rate of violence increased as the follow-up time increased both among civil psychiatric patients and offenders.

Moreover, every violent act committed by all participants in the near phase, far phase, and over the long-term follow-up periods was recorded by RAs (for the types, number of violent incidents, and time to first violent act across different sub-groups and time lengths, see Table S4).

#### 3.3 | Interrater reliability

Three RAs independently assessed 33 patients using the HCR-20<sup>V3</sup>, and the interrater reliability scores of these RAs was tested using the ICC, which was calculated using a two-way random model and the absolute agreement type based on the method of McGraw and Wong (1996). Table 2 displays the results, which indicate that the ICC of each index was above the moderate level<sup>1</sup> for single measures with a range of 0.58–0.93, and the ICC of each index was above the good level for average measures with a range of 0.80–0.98.

#### 3.4 | Concurrent validity

The correlation analysis between the sum of risk factors (includes presence rating and relevance rating) for the HCR-20 Total Scale, the SRRs, and the VRS Total was conducted. The Pearson's correlation coefficients between the



TABLE 2 ICC results associated with the HCR-20<sup>V3</sup>.

Index	Single measures	Average measures
	ICC (95% CI)	ICC (95% CI)
Σ Presence		
H	0.90 (0.83–0.94)	0.96 (0.93–0.98)
C	0.63 (0.44–0.78)	0.83 (0.70–0.91)
R	0.69 (0.52–0.82)	0.87 (0.76–0.93)
Total	0.82 (0.70–0.90)	0.93 (0.88–0.96)
Σ Relevance		
H	0.93 (0.88–0.96)	0.98 (0.96–0.99)
C	0.72 (0.57–0.84)	0.89 (0.80–0.94)
R	0.58 (0.36–0.75)	0.80 (0.63–0.90)
Total	0.88 (0.78–0.94)	0.96 (0.91–0.98)
SRRs	0.90 (0.82–0.94)	0.96 (0.93–0.98)

Abbreviations: HCR-20<sup>V3</sup>, Historical-Clinical-Risk Management-20 Version 3; SRR, summary risk rating.

VRS Total and HCR-20 Total Scale presence rating sum, relevance rating sum, and SRRs were 0.53, 0.71, and 0.64, respectively (all  $p$ s < 0.01), indicating good concurrent validity.

### 3.5 | Predictive validity

Tables 3 and 4 present the point biserial correlations ( $r_{pb}$ ) of and areas under the curve (AUC) between HCR-20<sup>V3</sup> indices and violence within 6 weeks and 6 months, which represent the predictive validity. In addition, Table S5 presents the  $r_{pb}$  and AUC between HCR-20<sup>V3</sup> indices and violence within 7–24 weeks (see Table S5). These results show that all SRRs, the HCR-20 Total Scale presence rating sums, and relevance rating sums were significantly associated with violence, and all AUCs were significant, regardless of sub-sample, gender, or follow-up length. Table S6 shows the means and standard deviations of correlations and AUC values within the three follow-up period lengths on the HCR-20<sup>V3</sup> indices; some independent-sample  $t$ -tests between civil psychiatric and offender sub-samples, males and females, were also performed on the means of each index (see Table S6).

Correlations between SRRs and violence ranged from 0.41 to 0.59 and AUC values ranged from 0.77 to 0.82. These results show that all AUCs of the SRRs met the standard of “excellent.”<sup>2</sup> Correlations between the HCR-20 Total Scale presence rating sums and violence ranged from 0.28 to 0.51 and AUC values ranged from 0.67 to 0.85. Based on the above criteria, all AUCs of the HCR-20 Total Scale presence rating sums met the standard of “adequate” or above. Correlations between the HCR-20 Total Scale relevance rating sums and violence ranged from 0.35 to 0.48 and AUC values ranged from 0.72 to 0.82. All AUCs of HCR-20 Total Scale relevance rating sums met the standard of “excellent.” These results indicate that across different follow-up lengths, the SRRs, the HCR-20 Total Scale presence rating sum, and the HCR-20 Total Scale relevance rating sum have satisfactory predictive validity for violence across different groups.

Within the combined sample, 3.2% (3 of 93) of people rated Low risk were violent during 6 weeks post-baseline interview; by comparison, 32.5% (13 of 40) of people in the Moderate-risk group conducted violent acts, and 50% (9 of 18) of people in the High-risk group acted violent incidents ( $X^2_{(2)} = 33.90$ ,  $p < 0.001$ ; Cramer's  $V = 0.47$ , 95% CI = [0.34, 0.62],  $p < 0.001$ ; Contingency Coefficient = 0.43, 95% CI = [0.32, 0.53],  $p < 0.001$ ). During the period of 7–24 weeks, 6.5% (6 of 92) of people in the Low-risk group exhibited violence, compared to 41% (16 of 39) of people rated Moderate and 61.1% (11 of 18) of people rated High ( $X^2_{(2)} = 36.94$ ,  $p < 0.001$ ; Cramer's  $V = 0.50$ , 95%

TABLE 3 The predictive validity results of the HCR-20<sup>V3</sup> indices on violence across different sub-groups within 6 weeks.

Violence within 6 weeks													
Civil		Offender			Male			Female			Combined		
Index	$r_{pb}$	95% CI	AUC	95% CI	$r_{pb}$	95% CI	AUC	95% CI	$r_{pb}$	95% CI	AUC	95% CI	
$\Sigma$ Presence													
H	0.36***	[0.16, 0.53]	0.77**	[0.62, 0.92]	0.31*	[0.08, 0.54]	0.74**	[0.60, 0.89]	0.35***	[0.17, 0.52]	0.78***	[0.66, 0.90]	
C	0.23*	[0.07, 0.37]	0.72*	[0.58, 0.86]	0.13[-0.18, 0.42]		0.58	[0.39, 0.78]	0.06[-0.15, 0.25]		0.55	[0.39, 0.71]	
R	0.30**	[0.10, 0.49]	0.73**	[0.61, 0.86]	0.28* [0.06, 0.50]		0.70*	[0.55, 0.85]	0.29** [0.09, 0.47]		0.72**	[0.59, 0.84]	
HCR-20 <sup>V3</sup>	0.44**	[0.24, 0.59]	0.84***	[0.72, 0.97]	0.34** [0.12, 0.55]		0.75**	[0.61, 0.89]	0.35*** [0.16, 0.52]		0.76***	[0.64, 0.89]	
$\Sigma$ Relevance													
H	0.40***	[0.19, 0.57]	0.80***	[0.66, 0.94]	0.27* [0.02, 0.51]		0.69*	[0.52, 0.86]	0.36*** [0.14, 0.54]		0.76***	[0.63, 0.90]	
C	0.18 [-0.001, 0.35]		0.68* [0.54, 0.83]		0.32* [0.02, 0.58]		0.71* [0.54, 0.88]		0.17 [-0.04, 0.39]		0.65 [0.50, 0.80]		
R	0.25* [0.07, 0.47]		0.73** [0.58, 0.88]		0.41*** [0.12, 0.63]		0.74** [0.58, 0.90]		0.39*** [0.16, 0.58]		0.77*** [0.63, 0.90]		
HCR-20 <sup>V3</sup>	0.35***	[0.14, 0.53]	0.80***	[0.65, 0.94]	0.39** [0.11, 0.60]		0.76**	[0.62, 0.90]	0.39*** [0.18, 0.56]		0.79***	[0.67, 0.90]	
SRR	0.48***	[0.26, 0.67]	0.82***	[0.69, 0.95]	0.45*** [0.22, 0.67]		0.81***	[0.68, 0.93]	0.43*** [0.25, 0.61]		0.81***	[0.71, 0.92]	
										0.54***	[0.23, 0.77]		
										0.47***	[0.31, 0.61]		

Abbreviations: AUC, areas under the curve; HCR-20<sup>V3</sup>, Historical-Clinical-Risk Management-20 Version 3; SRR, summary risk rating.  
\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; \* $p < 0.06$ .

TABLE 4 The predictive validity results of the HCR-20<sup>V3</sup> indices on violence across different sub-groups within 6 months.

Violence within 6 months																
Index	Civil			Offender			Male			Female			Combined			
	$r_{pb}$	95% CI	AUC	95% CI	$r_{pb}$	95% CI	AUC	95% CI	$r_{pb}$	95% CI	AUC	95% CI	$r_{pb}$	95% CI	AUC	95% CI
$\Sigma$ Presence																
H	0.45***	[0.28, 0.59]	0.78***	[0.68, 0.89]	0.20	[-0.04, 0.44]	0.65	[0.51, 0.79]	0.31**	[0.12, 0.48]	0.71***	[0.61, 0.82]	0.46***	[0.23, 0.64]	0.79***	[0.66, 0.92]
C	0.25*	[0.07, 0.42]	0.67*	[0.54, 0.79]	0.14	[-0.12, 0.38]	0.59	[0.43, 0.74]	0.12	[-0.09, 0.30]	0.57	[0.45, 0.70]	0.31*	[0.10, 0.49]	0.73**	[0.58, 0.89]
R	0.31**	[0.09, 0.51]	0.68**	[0.56, 0.81]	0.30*	[0.05, 0.53]	0.69*	[0.55, 0.83]	0.29**	[0.11, 0.48]	0.68**	[0.57, 0.80]	0.30*	[-0.01, 0.54]	0.67*	[0.50, 0.84]
HCR-20 <sup>v3</sup>	0.51***	[0.35, 0.65]	0.83***	[0.73, 0.92]	0.28*	[0.05, 0.50]	0.67*	[0.53, 0.81]	0.35***	[0.17, 0.51]	0.71**	[0.60, 0.81]	0.51***	[0.29, 0.68]	0.84***	[0.73, 0.95]
$\Sigma$ Relevance																
H	0.46***	[0.29, 0.62]	0.79***	[0.68, 0.90]	0.21	[-0.03, 0.45]	0.65	[0.50, 0.79]	0.35***	[0.17, 0.52]	0.74***	[0.63, 0.84]	0.42***	[0.19, 0.64]	0.76***	[0.61, 0.91]
C	0.34***	[0.14, 0.52]	0.73***	[0.62, 0.84]	0.30*	[-0.001, 0.56]	0.66a	[0.50, 0.82]	0.25*	[0.04, 0.46]	0.64*	[0.52, 0.77]	0.37**	[0.13, 0.62]	0.76**	[0.62, 0.90]
R	0.36***	[0.16, 0.57]	0.76***	[0.64, 0.87]	0.42***	[0.16, 0.61]	0.74**	[0.60, 0.88]	0.49***	[0.31, 0.64]	0.78***	[0.67, 0.88]	0.22	[-0.02, 0.50]	0.69*	[0.53, 0.84]
HCR-20 <sup>v3</sup>	0.48***	[0.29, 0.63]	0.81***	[0.71, 0.91]	0.35**	[0.11, 0.55]	0.72**	[0.59, 0.85]	0.44***	[0.26, 0.58]	0.78***	[0.68, 0.88]	0.41**	[0.19, 0.63]	0.76**	[0.62, 0.90]
SRR	0.59***	[0.39, 0.74]	0.82***	[0.71, 0.93]	0.46***	[0.23, 0.70]	0.77***	[0.64, 0.90]	0.55***	[0.38, 0.71]	0.81***	[0.71, 0.91]	0.50***	[0.21, 0.73]	0.78**	[0.63, 0.93]
															0.80***	[0.72, 0.88]

Abbreviations: AUC, areas under the curve; HCR-20<sup>V3</sup>, Historical-Clinical-Risk Management-20 Version 3; SRR, summary risk rating.

\**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001; <sup>a</sup>*p* < 0.06.

CI = [0.37, 0.65],  $p < 0.001$ ; Contingency Coefficient = 0.45; 95% CI = [0.35, 0.55],  $p < 0.001$ ). For the 6 months follow-up, 9.8% (9 of 92) of people rated Low risk conducted violent acts; during the same follow-up period, 53.8% (21 of 39) of people in the Moderate-risk group and 72.2% (13 of 18) of people in the High-risk group had violent acts ( $X^2_{(2)} = 44.65$ ,  $p < 0.001$ ; Cramer's  $V = 0.55$ , 95% CI = [0.42, 0.69],  $p < 0.001$ ; Contingency Coefficient = 0.48; 95% CI = [0.39, 0.57],  $p < 0.001$ ). Hence, within the combined sample, the rater-generated SRRs were significantly associated with violence, whether in the near phase, far phase, or over the long term.

The above results indicate that the main indices of the HCR-20<sup>V3</sup> are significantly associated with violence among patients with mental disorders over three follow-up lengths. In addition, we also found that sub-sample and gender themselves did not have moderating or confounding effects on the associations between the main indices of HCR-20<sup>V3</sup> and violence over three follow-up lengths (see detailed results in the Supporting Information).

### 3.6 | Examining incremental predictive validity of relevance ratings vis-à-vis presence ratings

Although it is not necessary for the validation of the HCR-20<sup>V3</sup> (Douglas et al., 2013), it could be the case that individual relevance ratings of risk factors possess incremental validity over and above the presence ratings of the same risk factors. First, the correlations between Presence and Relevance ratings (all  $ps < 0.001$ ) were 0.71 (H Scale), 0.65 (C Scale), 0.61 (R Scale), and 0.72 (Total Scale). Then, we examined whether Relevance ratings had stronger predictive power for violence than Presence ratings during the three follow-up periods by conducting 12 logistic regression analyses. Specifically, since we had verified that both sub-sample and gender had no moderating effect, we used the full sample for the analyses. Each of these analyses entered the sum of the Presence ratings (i.e., for the H scale) on step 1 and entered the corresponding sum of the Relevance ratings on step 2 (please see the results for the 7–24 weeks follow-up in the Supporting Information).

The results showed that, for the 6-week follow-up, the overall model for H Scale (Model  $X^2 = 22.16$ ,  $p < 0.001$ ,  $-2LL = 113.37$ , Nagelkerke  $R^2 = 0.23$ ), R Scale (Model  $X^2 = 15.17$ ,  $p = 0.001$ ,  $-2LL = 120.36$ , Nagelkerke  $R^2 = 0.16$ ), and HCR-20 Total Scale (Model  $X^2 = 26.38$ ,  $p < 0.001$ ,  $-2LL = 109.16$ , Nagelkerke  $R^2 = 0.27$ ) was significant, but the overall model for C Scale was only marginal significant (Model  $X^2 = 5.91$ ,  $p = 0.052$ ,  $-2LL = 129.63$ , Nagelkerke  $R^2 = 0.07$ ). The reason for this marginal significance of the overall model for C Scale was that their Presence ratings were not significantly associated with violence at 6 weeks ( $B = 0.19$ ,  $SE = 0.11$ , Wald = 2.80,  $p = 0.09$ ,  $\text{Exp}(B) = 1.21$ , 95% CI = [0.97, 1.51]; Model  $X^2 = 2.94$ ,  $p = 0.09$ ,  $-2LL = 132.59$ , Nagelkerke  $R^2 = 0.03$ ). Only for R Scale, Relevance ratings ( $B = 0.46$ ,  $SE = 0.24$ , Wald = 3.83,  $p = 0.05$ ,  $\text{Exp}(B) = 1.59$ , 95% CI = [1.00, 2.52]) added incrementally to Presence ratings (Block  $X^2 = 4.02$ ,  $p < 0.05$ ).

For the 6-month follow-up, the overall model for H Scale (Model  $X^2 = 24.82$ ,  $p < 0.001$ ,  $-2LL = 154.24$ , Nagelkerke  $R^2 = 0.22$ ), C Scale (Model  $X^2 = 12.63$ ,  $p < 0.01$ ,  $-2LL = 166.43$ , Nagelkerke  $R^2 = 0.12$ ), R Scale (Model  $X^2 = 21.96$ ,  $p < 0.001$ ,  $-2LL = 157.11$ , Nagelkerke  $R^2 = 0.20$ ), and HCR-20 Total Scale (Model  $X^2 = 32.49$ ,  $p < 0.001$ ,  $-2LL = 146.57$ , Nagelkerke  $R^2 = 0.28$ ) was significant. Relevance ratings of H Scale (Block  $X^2 = 4.97$ ,  $p = 0.03$ ), C Scale (Block  $X^2 = 8.98$ ,  $p = 0.003$ ), R Scale (Block  $X^2 = 9.38$ ,  $p = 0.002$ ), and HCR-20 Total Scale (Block  $X^2 = 5.48$ ,  $p = 0.02$ ) added incrementally to Presence ratings.

### 3.7 | Examining incremental predictive validity of summary risk ratings vis-à-vis presence ratings and relevance ratings

We separately conducted eight logistic regression analyses to examine whether SRRs had better predictive power for violence than either Presence Ratings or Relevance Ratings, for 6 weeks, 7–24 weeks, and 6 months periods. We used the full sample, given the absence of sub-sample moderating effects. Each of these eight analyses entered the Presence ratings or Relevance ratings index on step 1 and entered the SRRs on step 2.

For the 6-week follow-up, in the first Block of the model, all Presence (or Relevance) rating indices were significant, except for the C Scale Presence ratings. When we entered SRRs on step 2, SRRs were significant in each of the final models. Some of the Presence or Relevance ratings were no longer significant once the SRRs were entered into the model as follows: Risk Management Presence ratings ( $B = 0.20$ ,  $SE = 0.17$ ,  $Wald = 1.42$ ,  $p = 0.23$ ,  $Exp(B) = 1.22$ ,  $95\% CI = [0.88, 1.70]$ ), Historical Relevance ratings ( $B = 0.19$ ,  $SE = 0.12$ ,  $Wald = 2.38$ ,  $p = 0.12$ ,  $Exp(B) = 1.20$ ,  $95\% CI = [0.95, 1.52]$ ), Clinical Relevance ratings ( $B = -0.20$ ,  $SE = 0.18$ ,  $Wald = 1.31$ ,  $p = 0.25$ ,  $Exp(B) = 0.82$ ,  $95\% CI = [0.58, 1.16]$ ), Risk Management Relevance ratings ( $B = 0.01$ ,  $SE = 0.26$ ,  $Wald = 0.002$ ,  $p = 0.96$ ,  $Exp(B) = 1.01$ ,  $95\% CI = [0.61, 1.68]$ ), and Total Relevance ratings ( $B = 0.05$ ,  $SE = 0.08$ ,  $Wald = 0.34$ ,  $p = 0.56$ ,  $Exp(B) = 1.05$ ,  $95\% CI = [0.89, 1.23]$ ). The overall models of these eight analyses were significant.

With violence within 7–24 weeks as the outcome, all Presence (or Relevance) rating indices were significant in the first Block of the model. When SRRs were entered on step 2, all SRRs were significant in the final model, whereas all Presence (or Relevance) ratings index were no longer significant. Moreover, the overall models of these eight analyses were significant.

Conducting another eight regression analyses with violence at 6 months as the outcome, all Presence (or Relevance) ratings index were significant in the first Block of the model, except for the C Scale Presence ratings ( $B = 0.17$ ,  $SE = 0.09$ ,  $Wald = 3.50$ ,  $p = 0.06$ ,  $Exp(B) = 1.19$ ,  $95\% CI = [0.99, 1.42]$ ). When we entered SRRs on step 2, all SRRs were significant. The following Presence or Relevance ratings were no longer significant once SRRs were in the model: R Scale Presence ( $B = 0.14$ ,  $SE = 0.15$ ,  $Wald = 0.94$ ,  $p = 0.33$ ,  $Exp(B) = 1.15$ ,  $95\% CI = [0.86, 1.54]$ ), Total Scale Presence ( $B = 0.11$ ,  $SE = 0.06$ ,  $Wald = 2.87$ ,  $p = 0.09$ ,  $Exp(B) = 1.11$ ,  $95\% CI = [0.98, 1.26]$ ), H Scale Relevance ( $B = 0.12$ ,  $SE = 0.10$ ,  $Wald = 1.36$ ,  $p = 0.24$ ,  $Exp(B) = 1.13$ ,  $95\% CI = [0.92, 1.38]$ ), C Scale Relevance ( $B = -0.07$ ,  $SE = 0.15$ ,  $Wald = 0.22$ ,  $p = 0.64$ ,  $Exp(B) = 0.93$ ,  $95\% CI = [0.70, 1.25]$ ), R Scale Relevance ( $B = 0.11$ ,  $SE = 0.24$ ,  $Wald = 0.20$ ,  $p = 0.66$ ,  $Exp(B) = 1.11$ ,  $95\% CI = [0.69, 1.79]$ ), and Total Scale Relevance ( $B = 0.05$ ,  $SE = 0.07$ ,  $Wald = 0.54$ ,  $p = 0.46$ ,  $Exp(B) = 1.06$ ,  $95\% CI = [0.92, 1.22]$ ). The overall models of these eight analyses were significant.

Overall, these results basically indicated that the SRRs added incrementally to the sum of the presence or relevance of risk factors for the prediction of violence over three follow-up periods.

### 3.8 | Examining incremental predictive validity of HCR-20<sup>V3</sup> indices vis-à-vis VRS index

We examined whether HCR-20<sup>V3</sup> (and its components) had stronger predictive power for violence than VRS during the three follow-up periods by conducting nine logistic regression analyses. Each of these analyses entered the VRS Total (zero-centered) on step 1 and entered the HCR-20<sup>V3</sup> indices (zero-centered) on step 2, respectively.

The results in the Table 5 showed that, for the 6-week follow-up, the VRS Total was significant in the first Block of the model,  $B = 0.11$ ,  $Wald = 13.67$ ,  $p < 0.001$ ,  $Exp(B) = 1.12$ ,  $95\% CI = [1.05, 1.18]$ . When we entered the HCR-20<sup>V3</sup> indices, the HCR-20<sup>V3</sup> indices were significant in each of the final models. The VRS Total was no longer significant when Total Presence ratings, Total Relevance ratings, and the SRRs were entered, respectively. The overall models of these three analyses were significant.

With violence within 7–24 weeks as the outcome, the VRS Total was significant in the first Block of the model,  $B = 0.09$ ,  $Wald = 12.27$ ,  $p < 0.001$ ,  $Exp(B) = 1.10$ ,  $95\% CI = [1.04, 1.15]$ . When the HCR-20<sup>V3</sup> indices were entered on step 2, all HCR-20<sup>V3</sup> indices were significant in the final model, whereas all the VRS Total ratings were no longer significant. Moreover, the overall models of these three analyses were significant.

Conducting another three regression analyses with violence within 6 months as the outcome, the VRS Total was significant in the first Block of the model,  $B = 0.11$ ,  $Wald = 16.71$ ,  $p < 0.001$ ,  $Exp(B) = 1.11$ ,  $95\% CI = [1.06, 1.17]$ . When we entered the HCR-20<sup>V3</sup> indices on step 2, all HCR-20<sup>V3</sup> indices were significant. The VRS Total ratings were no longer significant when Total Relevance ratings and the SRRs were entered. When Total Presence ratings was entered, the VRS Total ratings were still significant ( $B = 0.07$ ,  $Wald = 5.46$ ,  $p = 0.02$ ,  $Exp(B) = 1.07$ ,  $95\% CI = [1.01, 1.14]$ ). The overall models of these three analyses were significant.

**TABLE 5** The incremental predictive validity results of HCR-20<sup>V3</sup> indices vis-à-vis VRS index across time lengths.

	<i>B</i> ( <i>SE</i> )	Exp( <i>B</i> ) [95% CI]	Wald	<i>p</i>
Violence within 6 weeks				
VRS-total	0.06 (0.04)	1.07 [1.00, 1.14]	3.29	0.07
P-total	0.21 (0.07)	1.24 [1.08, 1.42]	9.13	0.003
		$X^2(2) = 28.40, p < 0.001$		
VRS-total	0.06 (0.04)	1.06 [0.99, 1.15]	2.63	0.11
R-total	0.16 (0.07)	1.17 [1.02, 1.35]	4.82	0.03
		$X^2(2) = 22.86, p < 0.001$		
VRS-total	0.03 (0.04)	1.03 [0.96, 1.11]	0.82	0.37
HCR SRRs	1.42 (0.42)	4.12 [1.81, 9.36]	11.40	0.001
		$X^2(2) = 30.71, p < 0.001$		
Violence within 7–24 weeks				
VRS-total	0.05 (0.03)	1.05 [0.99, 1.11]	2.41	0.12
P-total	0.19 (0.06)	1.21 [1.07, 1.36]	9.46	0.002
		$X^2(2) = 25.61, p < 0.001$		
VRS-total	0.02 (0.04)	1.02 [0.95, 1.09]	0.35	0.55
R-total	0.22 (0.07)	1.25 [1.09, 1.43]	10.53	0.001
		$X^2(2) = 26.68, p < 0.001$		
VRS-total	0.002 (0.03)	1.00 [0.94, 1.07]	0.01	0.94
HCR SRRs	1.62 (0.41)	5.04 [2.26, 11.25]	15.60	0.000
		$X^2(2) = 33.75, p < 0.001$		
Violence within 6 months				
VRS-total	0.07 (0.03)	1.07 [1.01, 1.14]	5.46	0.02
P-total	0.18 (0.06)	1.20 [1.07, 1.34]	9.98	0.002
		$X^2(2) = 32.88, p < 0.001$		
VRS-total	0.05 (0.03)	1.05 [0.99, 1.12]	2.34	0.13
R-total	0.19 (0.06)	1.21 [1.07, 1.37]	8.81	0.003
		$X^2(2) = 31.18, p < 0.001$		
VRS-total	0.02 (0.03)	1.02 [0.96, 1.09]	0.37	0.54
HCR SRRs	1.63 (0.40)	5.10 [2.34, 11.12]	16.78	0.000
		$X^2(2) = 42.42, p < 0.001$		

Note: “P” indicates Presence Rating, and “R” indicates Relevance Rating. “P-Total” is the sum of Presence Ratings for the HCR-20<sup>V3</sup>, “R-Total” is the sum of Relevance Ratings for the HCR-20<sup>V3</sup>.

Abbreviations: AUC, areas under the curve; HCR-20<sup>V3</sup>, Historical-Clinical-Risk Management-20 Version 3; SRR, summary risk rating; VRS, Violence Risk Scale.

In general, these results basically indicated that the HCR-20<sup>V3</sup> indices added incrementally to the VRS indices for the prediction of violence over three follow-up periods.

## 4 | DISCUSSION

### 4.1 | Participant characteristics and sample size

The HCR-20<sup>V3</sup> has been used by scholars worldwide, including countries such as Canada (e.g., Blanchard & Douglas, 2011; Hogan & Olver, 2016, 2019; Penney et al., 2016; Strub et al., 2014), Netherlands (e.g., de Vogel et al., 2014; de Vries Robbé & de Vogel, 2010), Norway (e.g., Eidhammer et al., 2013), Sweden (e.g., Douglas & Belfrage, 2014; Persson et al., 2017), United Kingdom (e.g., Doyle et al., 2013, 2014), United States (e.g., Green et al., 2016; Howe et al., 2016; Smith & Edens, 2013; Smith et al., 2014), Germany (e.g., Kötter et al., 2014), Ghana (Adjorlolo & Chan, 2019), South Korea (Sea & Bang, 2021), Mexico (Nijdam-Jones et al., 2021, 2022), and so on, to investigate and verify its reliabilities and validities in their countries. Combining the aforementioned 21 empirical studies, which contain 10 forensic samples, six civil samples, and seven correctional samples, the total sample size was 1942; the smallest sample was 20 (Eidhammer et al., 2013) and the largest was 387 (Doyle et al., 2014) with an average sample size of 97.

This study represents the first use of the HCR-20<sup>V3</sup> to simultaneously empirically investigate offenders with mental disorders and civil psychiatric patients in China. A total of 152 participants with mental disorders were assessed, which are above the average sample size of other studies. Moreover, Schönbrodt and Perugini (2013) indicated that the critical N for stability of course decreases as the effect size increases. In this study, we were dealing mainly with the equivalent of moderate to large correlations and beta coefficients (0.30–0.50); the sample size of 150+ is justifiable according to Schönbrodt and Perugini (2013).

### 4.2 | Reliability and validity of the Chinese translation of HCR-20<sup>V3</sup>

In the revision of Dutch translation of HCR-20<sup>V3</sup>, de Vogel et al. (2014) suggested that future validation studies should focus on interrater reliability, concurrent, and predictive validity for the assessment of violence risk in different settings and countries. The current study tested the interrater reliability, concurrent validity, and predictive validity of the Chinese translation of the HCR-20<sup>V3</sup>. Regarding interrater reliability, three raters independently evaluated 33 mentally disordered patients in this study. The sample sizes of previous interrater reliability analyses ranged from 10 (Persson et al., 2017) to 80 (Sea & Bang, 2021) with an average of 31. The results of our analyses showed that the ICC of Presence ratings was similar to that of the Relevance ratings, and both were somewhat lower than that of the SRRs. These ranges of single and average measure ICCs in this study were very similar to those obtained by Douglas and Belfrage's (2014) interrater reliability study, wherein three raters independently evaluated 35 forensic patients. Comparing with the results of another study conducted in East Asian country (Sea & Bang, 2021), we found that the values of five of the nine indices (i.e., the single measure ICC values of the H and Total Presence ratings, H and Total Relevance ratings, and SRRs) were greater in this study. Our findings indicate that the interrater reliability of the Chinese translation of the HCR-20<sup>V3</sup> is satisfactory.

Regarding concurrent validity, past studies have mostly used the HCR-20<sup>V2</sup> to calculate the correlations (5 of the 9 empirical studies); in terms of the presence rating sum of the Total Scale, the correlation coefficients between the HCR-20<sup>V3</sup> and HCR-20<sup>V2</sup> varied from 0.84 (Eidhammer et al., 2013) to 0.93 (de Vries Robbé & de Vogel, 2010) with an average of 0.88. Other tools have also been used to test the concurrent validity of the HCR-20<sup>V3</sup>; for example, the correlation coefficients of PCL:SV Total for the HCR-20<sup>V3</sup> Total Scale presence rating sum, relevance rating sum, and SRRs were 0.71, 0.79, and 0.82, respectively (Blanchard & Douglas, 2011); the correlation coefficients of the PCL-R Total for the HCR-20<sup>V3</sup> Total Scale presence rating sum (R-In) and (R-Out) were 0.66 and 0.70, respectively (Smith & Edens, 2013); the correlation coefficients of the HCR-20<sup>V3</sup> Total Scale presence rating sum with regard to COVR, LSI-R, and SAPROF Total were 0.43, 0.64, and −0.51, respectively (Persson et al., 2017); the correlation coefficients of the HCR-20<sup>V3</sup> Total Scale presence rating sum with regard to CAPP-IRS, PCL-R, LSI-R, and KORAS-G Total were



0.54, 0.44, 0.58, and 0.44, respectively (Sea & Bang, 2021); and the correlation coefficients of VRS Total for the HCR-20<sup>V3</sup> Total Scale presence rating sum, relevance rating sum, and SRRs were 0.77, 0.81, and 0.76, respectively (Hogan & Olver, 2016). Compared with and adding to previous studies by using a different comparative instrument, the concurrent validity of the Chinese translation of the HCR-20<sup>V3</sup> can be considered satisfactory.

Concerning predictive validity, the present study adopted a prospective design, which is in line with most of the past studies. In terms of the follow-up length for violence, we adopted two time-frames (6 weeks and 6 months) that were consistent with Strub et al. (2014). In this study, however, we innovatively treated the first follow-up visit (i.e., the sixth week post-baseline) as a time node and divided the 6-month period after the baseline interview into two phases (i.e., the “near phase” and “far phase”) to facilitate further examination of the predictive validity of the main indices of the HCR-20<sup>V3</sup>.

We calculated the  $r_{pb}$  and AUCs values of the SRRs for violence by sub-sample (i.e., combined, hospital patients, and offenders) within 6 weeks and 6 months, respectively. Compared with the results of Strub et al. (2014), the values of nine of the 12 aforementioned indices in this study are greater. By comparing the data recorded 6 months after the baseline interview of this study with those of recorded within 6 months by Doyle et al. (2014), we found that the values of six of the eight indices (i.e., the  $r_{pb}$  and AUCs values of the H, C, R, and Total Presence ratings for violent acts) were greater in this study than in Doyle et al. (2014). As such, compared to these Canadian and UK samples of mentally disordered offenders, civil psychiatric patients, and forensic psychiatric patients, the HCR-20<sup>V3</sup> is performing at least as well with Chinese mentally disordered offenders and civil psychiatric patients in terms of basic predictive validity.

Furthermore, we found that the HCR-20<sup>V3</sup> Relevance ratings tended to add incremental predictive validity to the presence ratings. This finding makes conceptual sense, in that risk factors that are rated not only as present, but as relevant (i.e., deemed to be causally connected at the individual level to violence for the given individual), ought to be more strongly related to violence. Similar findings were reported by Hogan and Olver (2016, 2019) in two Canadian studies of forensic psychiatric patients.

Finally, SRRs tended to increment presence and relevance ratings, a finding that is consistent with past research on the HCR-20<sup>V3</sup> (for a review, see Douglas & Shaffer, 2021) as well as numerous studies using Version 2 (for reviews, see Douglas & Reeves, 2010; Guy et al., 2015; Heilbrun et al., 2021) and other Structured Professional Judgment (SPJ) measures. The final ratings of SPJ measures are made to indicate low, moderate, or high risk for violence. To be clear, this is an example of structured human judgment outperforming algorithmic prediction, in the form of sums of risk factor ratings being statistically optimized within a logistic regression model. Although this is used to be somewhat of a controversial finding years ago, as some commentators argued that it was inconsistent with Meehl's (1954) proclamation of the superiority of actuarial prediction over clinical prediction, the robustness of this finding across SPJ instruments, across time, across place, and across translations is very strong (Garb & Wood, 2019; Guy et al., 2015; Heilbrun et al., 2021). Although the way this instrument is used in clinical practice does not directly match the way these data are analyzed, that is, users will not calculate the total scores for the H, C, R, and Total scale, we want to remind users that the number of risk factors in a given case is, in fact, important. As stated in the HCR-20<sup>V3</sup> user guide (see p. 62–63): “Generally, the more risk factors that are present and relevant, the higher the risk for future violence. Although SRRs of low, moderate and high risk always require justification by evaluators, we encourage evaluators who make decisions of high risk in the face of few risk factors, or low risk in the face of many, to explain why their decision does not comport with the general rule of ‘more risk factors, higher risk’.”

We believed that it was necessary to investigate the effect of sub-sample type and gender on the predictive power of the main indices of the HCR-20<sup>V3</sup> over different time lengths among Chinese patients. The descriptive comparisons of this paper showed that the predictive powers of the HCR-20<sup>V3</sup> indices with regard to violence showed certain commonalities and variations between the hospital patient and offender sub-samples and between male and female patients. As such, we sought to verify whether sub-sample type and gender significantly affected the predictive power of the HCR-20<sup>V3</sup> indices using approaches with more statistical power. The results of a series of logistic regression analyses in this study showed that across three time lengths, sub-sample type and gender did



not moderate or confound the predictive power of the main indices of the HCR-20<sup>V3</sup> among Chinese patients. These findings are consistent with the findings of Strub et al. (2014) in Canada (with respect to gender and sub-sample) as well as with the conclusion of Green et al. (2016) in the United States that gender did not moderate the correlation between violence and the HCR-20<sup>V3</sup> indices.

Together, these results indicate that the predictive and incremental validity of the main indices of the Chinese translation of the HCR-20<sup>V3</sup> are satisfactory and are robust across both gender and setting (psychiatric patients; mentally disordered offenders). As HCR-20<sup>V3</sup> is one of the SPJ measures, the authors of HCR-20<sup>V3</sup> have repeatedly stressed that its application is a non-algorithmic and non-numeric decision process. It helps avoid the pitfalls inherent in actuarial approaches, such as sample dependence, exclusion of potentially important risk factors, instability of precise probability estimates across samples, and the inherent difficulty in applying group-based probability estimates to individuals (Douglas et al., 2014). It is indicated that the HCR-20<sup>V3</sup> attaches great importance to the professional judgment of the rater, while emphasizing the uniqueness of each participant. As such, the finding that the SRRs were rated reliably, possessed strong predictive validity, and added incremental validity to the Presence and Relevance ratings of the risk factors provides strong support for one of the core assumptions of the SPJ model generally and the HCR-20<sup>V3</sup> specifically.

To date, although the psychometric properties of the HCR-20<sup>V3</sup> have been tested in many populations, most of the studies with the HCR-20<sup>V3</sup> were conducted in Western countries. Obviously, more research on the psychometric properties of the HCR-20<sup>V3</sup> in Non-Western populations is needed. Zhou et al. (2016) reviewed risk assessment studies in China and concluded that there is limited evidence to support the use of violence risk assessment instruments in general and forensic patients in China. This study proves that HCR-20<sup>V3</sup> has good reliability and validity in assessing the risk of violence in patients with mental disorders in the Chinese population.

### 4.3 | Limitations and future directions

Although this study had several strengths (e.g., described the first use of the HCR-20<sup>V3</sup> in China with a moderately sized sample of participants incarcerated or hospitalized; prospective design), some limitations must be noted as well.

First, the offender sub-sample lacked female participants, which led to a mismatch in the characteristics of the people in the two sub-samples and affected the balances between the numbers of male and female participants as well as the balance between the hospital patient and offender sub-samples. Thus, the statistical power of the results of this study was somewhat weakened.

Second, we applied convenience sampling by only choosing female patients from the same ward of the same psychiatric hospital when testing interrater reliability. According to Green et al. (2016), male participants obtained better interrater reliability values than females, and future studies should investigate the potential effect of participants' gender on interrater reliability tests. Therefore, our future studies will verify the interrater reliability of the Chinese version the HCR-20<sup>V3</sup> among male and female participants separately.

Third, this study was limited by objective conditions, such as time, and we only validated the predictive power of the HCR-20<sup>V3</sup> with regard to the violent acts of mentally disordered patients over six weeks and six months. We will continuously follow up with the same sample so that data regarding the predictive power of violent acts are obtained for longer follow-up lengths.

Fourth, this study used the traditional statistical methods to analyze the data, because it is easier to compare the results with former studies from a cross-cultural perspective. We will conduct analyses based on time-to-failure using survival analysis in the following studies. Besides, this study took an etic approach to validate the HCR-20<sup>V3</sup> in Chinese population. By contrast, an emic approach would have evaluated the appropriateness of the HCR-20<sup>V3</sup> administration procedure and risk factors. Future studies could take an emic approach to interview some clinicians or other key stakeholders about the appropriateness of what is currently included in the HCR-20<sup>V3</sup> guidelines, as well as whether there is anything that ought to be added to them.

Fifth, in this study, two raters interviewed one participant at the same time, which is systematically different from Europe and North America's procedure. Because a part of our participants came from the prison, two or more raters had been recommended to participate in an interview room for security and human rights perspectives. In the administration, one rater conducted the interview and the other rater observed. Then, the two raters independently rated each item of the HCR-20<sup>V3</sup> based on their interview notes and the information collected. Nevertheless, as two raters in a same room could possibly influence each other, it is needed to test whether the number of raters might have an impact on the ratings.

The present study aimed to contribute to the literature on risk assessment from a Non-Western context, specifically China, an East Asian nation. We believe that China and other Non-Western countries can also benefit from risk assessment practices. Risk assessment could possibly contribute significantly to justice delivery, offender rehabilitation, and ultimately public safety (Adjorlolo & Chan, 2019). We will continue to test the psychometric properties of the HCR-20<sup>V3</sup> among non-clinical samples in China. Besides, researchers indicated that cross-cultural characteristics (e.g., collectivism and patriarchal social system) may play an important role in the risk representation (Sea & Bang, 2021). We will compare the results of HCR-20<sup>V3</sup> with some other Chinese-originated tools to discuss the cultural factors in violence risk assessment in the future.

## 5 | CONCLUSION

This paper describes the first use of the HCR-20<sup>V3</sup> in China. Risk assessments of violence were made for (1) prison inmates with mental disorders and (2) civil psychiatric patients. The interrater reliability, concurrent validity, and predictive validity of the HCR-20<sup>V3</sup> are described. Our findings indicate that the interrater reliability and concurrent validity of the Chinese translation of the HCR-20<sup>V3</sup> are satisfactory. By adopting a prospective design and comparing to the results previously found in other countries' sample of mentally disordered offenders, civil psychiatric patients, and forensic psychiatric patients, the HCR-20<sup>V3</sup> is performing at least as well with Chinese mentally disordered offenders and civil psychiatric patients in terms of basic predictive validity.

Our findings also indicate that most individuals with psychiatric disorders are not aggressive. It can be seen that the number of patients with serious mental disorders is increasing year by year at a high rate in China. However, there is limited work in Chinese populations with psychiatric disorders using violence risk assessment tools. We believe that the Chinese translation of the HCR-20<sup>V3</sup> can help to find the relatively small cohort of individuals with a high risk. As such, this measurement instrument may help not only to protect the personal safety of the public but also to remove the stigma that the general public has of mentally ill persons.

## ACKNOWLEDGMENTS

Supported by the Fundamental Research Funds for the Central Universities, Grant Number: 2022JKF427 and 202ZFD82002.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## ENDNOTES

<sup>1</sup> According to Landis and Koch (1977), the ICC value should be greater than 0.80 (0.61–0.80: good; 0.41–0.60: moderate; 0.11–0.40: low; below 0.1: no consistency).

<sup>2</sup> Rice and Harris (2005) recommended that the AUC values are "acceptable" (over 0.556), "adequate" (over 0.639), and "excellent" (over 0.714).

## REFERENCES

- Adjorlolo, S., & Chan, H. C. O. (2019). Risk assessment of criminal offenders in Ghana: An investigation of the discriminant validity of the HCR-20<sup>V3</sup>. *International Journal of Law and Psychiatry*, 66, 101458. <https://doi.org/10.1016/j.ijlp.2019.101458>
- Andrews, D. A., & Bonta, J. (1994). *The psychology of criminal conduct*. Anderson Publishing Co.
- Andrews, D. A., Bonta, J. A., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency*, 52(1), 7–27. <https://doi.org/10.1177/0011128705281756>
- Blackburn, R. (2004). "What works" with mentally disordered offenders. *Psychology, Crime and Law*, 10(3), 297–308. <https://doi.org/10.1080/10683160410001662780>
- Blanchard, A., & Douglas, K. S. (2011). The historical-clinical-risk management-version 3: The inclusion of idiographic relevance ratings in violence risk assessment. In *Invited poster presented at the annual convention of the American-Psychology Law Society, Miami, FL*.
- Bonta, J., Blais, J., & Wilson, H. A. (2014). A theoretically informed meta-analysis of the risk for general and violent recidivism for mentally disordered offenders. *Aggression and Violent Behavior*, 19(3), 278–287. <https://doi.org/10.1016/j.avb.2014.04.014>
- de Vogel, V., De Beuf, T., Shepherd, S., & Schneider, R. D. (2022). Violence risk assessment with the HCR-20<sup>V3</sup> in legal contexts: A critical reflection. *Journal of Personality Assessment*, 104(2), 252–264. <https://doi.org/10.1080/00223891.2021.2021925>
- 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. <https://doi.org/10.1080/14999013.2014.906518>
- de Vries Robbé, M., & de Vogel, V. (2010). *Pilot study of HCR-20 (version 3)*. Van der Hoeven Kliniek. Unpublished data analyses.
- Douglas, K. S. (2019). Mental illness and violent behavior: Perceptions and evidence. In R. D. Morgan (Ed.), *Sage encyclopedia of criminal psychology*. Sage.
- 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. <https://doi.org/10.1080/14999013.2014.908429>
- Douglas, K. S., Guy, L. S., & Hart, S. D. (2009). Psychosis as a risk factor for violence to others: A meta-analysis. *Psychological Bulletin*, 135(5), 679–706. <https://doi.org/10.1037/a0016311>
- Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20V3: Professional guidelines for evaluating risk of violence*. 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-20<sup>V3</sup>): Development and overview. *International Journal of Forensic Mental Health*, 13(2), 93–108. <https://doi.org/10.1080/14999013.2014.906519>
- Douglas, K. S., & Reeves, K. (2010). The HCR-20 violence risk assessment scheme: Overview and review of the research. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment*. Routledge/Taylor and Francis.
- Douglas, K. S., & Shaffer, C. S. (2021). The science of and practice with the HCR-20V3 (Historical-Clinical-Risk management-20, version 3). In K. S. Douglas & R. K. Otto (Eds.), *Handbook of violence risk assessment* (2nd ed., pp. 253–293). Routledge/Taylor and Francis.
- 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-20<sup>V3</sup>. *International Journal of Forensic Mental Health*, 13(2), 140–147. <https://doi.org/10.1080/14999013.2014.906517>
- Doyle, M., Shaw, J., & Coid, J. (2013). *Validation of risk assessments in medium secure services (VORAMSS): Preliminary HCR-20 (Version 3) analyses*. University of Manchester, UK, and Queen Mary University of London, UK. Unpublished data analyses.
- Eidhammer, G., Selmer, L. E., Bjørkly, S., & SAFE Pilot Project Group. (2013). *A brief report of a comparison of the HCR-20 (Version 2) and HCR-20 (Version 3) concerning internal consistency and clinical utility*. Centre for Research and Education in Forensic Psychiatry, Oslo University Hospital.
- Gao, S. Y., & Fei, L. P. (2001). Attitudes about mental illness of different types of respondents in Beijing. *Chinese Mental Health Journal*, 2, 107–109.
- Garb, H. N., & Wood, J. M. (2019). Methodological advances in statistical prediction. *Psychological Assessment*, 31(12), 1456–1466. <https://doi.org/10.1037/pas0000673>
- Green, D., Schneider, M., Griswold, H., Belfi, B., Herrera, M., & Deblasi, A. (2016). A comparison of the HCR-20<sup>V3</sup> among male and female insanity acquittees: A retrospective file study. *International Journal of Forensic Mental Health*, 15(1), 48–64. <https://doi.org/10.1080/14999013.2015.1134726>
- Gu, Y., Singh, J. P., Yun, L. B., & Hu, Z. Q. (2014). A review of violence risk assessment for mentally disordered patients in mainland of China. *Criminal Justice and Behavior*, 41(12), 1398–1405. <https://doi.org/10.1177/0093854814547950>
- Guy, L. S., & Douglas, K. S. (2015). Major mental disorders and violence. In C. A. Pietz & C. A. Mattson (Eds.), *Violent offenders: Understanding and assessment* (pp. 77–98). Oxford University Press.

- Guy, L. S., Douglas, K. S., & Hart, S. D. (2015). Risk assessment and communication. In B. Cutler & P. Zapf (Eds.), *APA handbook of forensic psychology* (Vol. 1, pp. 35–86). American Psychological Association.
- Heilbrun, K., Hart, S. D., Hare, R. D., Gustafson, D., Nunez, C., & White, A. J. (1998). Inpatient and postdischarge aggression in mentally disordered offenders: The role of psychopathy. *Journal of Interpersonal Violence*, 13(4), 514–527. <https://doi.org/10.1177/088626098013004007>
- Heilbrun, K., Yasuhara, K., Shah, S., & Locklair, B. (2021). Approaches to violence risk assessment: Overview, critical analysis, and future directions. In K. S. Douglas & R. K. Otto (Eds.), *Handbook of violence risk assessment* (2nd ed., pp. 3–27). Routledge/Taylor & Francis.
- 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. <https://doi.org/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. <https://doi.org/10.1177/0093854819846526>
- 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. <https://doi.org/10.1177/0093854815605527>
- Huang, Y. Q., Wang, Y., Wang, H., Liu, Z. R., Yu, X., Yan, J., Yu Y., Kou C., Xu X., Lu J., Wang Z., He S., Xu Y., He Y., Li T., Guo W., Tian H., Xu G., Xu X., ... Wu, Y. (2019). Prevalence of mental disorders in China: A cross-sectional epidemiological study. *Lancet Psychiatry*, 6(3), 211–224. [https://doi.org/10.1016/S2215-0366\(18\)30511-X](https://doi.org/10.1016/S2215-0366(18)30511-X)
- Kötter, S., Von Franqué, F., Bolzmacher, M., Eucker, S., Holzinger, B., & Müller-Isberner, R. (2014). The HCR-20<sup>V3</sup> in Germany. *International Journal of Forensic Mental Health*, 13(2), 122–129. <https://doi.org/10.1080/14999013.2014.911784>
- Kropp, P. R., Hart, S. D., & Lyon, D. (2008). *Guidelines for stalking assessment and management (SAM): User manual*. ProActive Resolutions Inc.
- Landis, J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46. <https://doi.org/10.1037/1082-989X.1.4.390>
- Meehl, P. E. (1954). *Clinical vs. statistical prediction*. University of Minnesota Press.
- Monahan, J. (1992). "A terror to their neighbors": Beliefs about mental disorder and violence in historical and cultural perspective. *Bulletin of the American Academy of Psychiatry and Law*, 20(2), 191–195.
- Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., Roth, L. H., Grisso, T., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur Study of mental disorder and violence*. Oxford University Press.
- Nijdam-Jones, A., García-López, E., Merchan-Rojas, L., Guarneros, A. R., & Rosenfeld, B. (2021). Predictive validity of the HCR-20<sup>V3</sup> with incarcerated males in Mexico City. *Criminal Justice and Behavior*, 48(10), 1450–1467. <https://doi.org/10.1177/0093854821997520>
- Nijdam-Jones, A., García-López, E., Merchan-Rojas, L., Guarneros, A. R., & Rosenfeld, B. (2022). Cross-cultural violence risk assessment: Adapting the HCR-20<sup>V3</sup> for incarcerated offenders in Mexico. *International Journal of Forensic Mental Health*, 22(1), 39–55. <https://doi.org/10.1080/14999013.2022.2053000>
- 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. <https://doi.org/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<sup>V3</sup>, and SAPROF. *International Journal of Forensic Mental Health*, 16(2), 117–129. <https://doi.org/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. <https://doi.org/10.1007/s10979-005-6832-7>
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5), 609–612. <https://doi.org/10.1016/j.jrp.2013.05.009>
- Sea, J., & Bang, S. (2021). The interrater reliability and concurrent validity of the HCR-20 version 3 in South Korea. *Journal of Forensic Psychiatry and Psychology*, 32(6), 879–901. <https://doi.org/10.1080/14789949.2021.1937279>
- Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach, 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>
- Smith, S., & Edens, J. F. (2013). *Concurrent validity between the HCR-20 (version 3) and the PCL-R among male pre-trial offenders in Texas*. Texas A&M University. Unpublished data analyses.
- Smith, S. T., Kelley, S. E., Rulseh, A., Sörman, k., & Edens, J. F. (2014). Adapting the HCR-20<sup>V3</sup> for pre-trial settings. *International Journal of Forensic Mental Health*, 13(2), 160–171. <https://doi.org/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. *International Journal of Forensic Mental Health*, 13(2), 148–159. <https://doi.org/10.1080/14999013.2014.911785>
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence* (Version 2). Mental Health, Law, and Policy Institute, Simon Fraser University.
- Webster, C. D., Douglas, K. S., & Wintrup, A. (1995). *The HCR-20 scheme: The assessment of dangerousness and risk*. Mental Health, Law, and Policy Institute, Simon Fraser University.
- Wong, S., & Gordon, A. E. (1999–2003). *The violence risk scale*. University of Saskatchewan. Unpublished manuscript.
- Zhang, X. L., Chen, X. C., Cai, W. X., & Hu, J. M. (2012). Reliability of the violence risk scale of Chinese version. *Journal of Forensic Medicine*, 28(1), 32–35. <https://doi.org/10.3969/j.issn.1004-5619.2012.01.008>
- Zhou, J. S., Witt, K., Xiang, Y. T., Zhu, X. M., Wang, X. P., & Fazel, S. (2016). Violence risk assessment in psychiatric patients in China: A systematic review. *Australian and New Zealand Journal of Psychiatry*, 50(1), 33–45. <https://doi.org/10.1177/0004867415585580>

## SUPPORTING INFORMATION

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

**How to cite this article:** Chen, Y., Douglas, K. S., Zhang, Z., Xiao, C., Wang, H., Wang, Y., & Ma, A. (2023). Evaluating the HCR-20<sup>V3</sup> violence risk assessment measure with mentally disordered offenders and civil psychiatric patients in China. *Behavioral Sciences & the Law*, 41(4), 186–206. <https://doi.org/10.1002/bsl.2614>