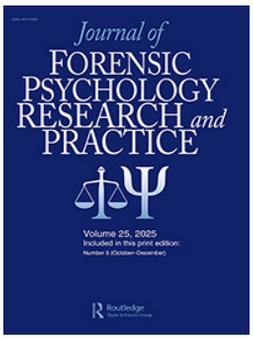

Routledge
Taylor & Francis Group

Journal of Forensic Psychology Research and Practice

ISSN: 2473-2850 (Print) 2473-2842 (Online) Journal homepage: www.tandfonline.com/journals/wfpp21

Use of the HCR-20 Version 3 with Women: A Narrative Synthesis

Cevher Gunenc, Lucy Fitton & Peter Beazley

To cite this article: Cevher Gunenc, Lucy Fitton & Peter Beazley (2025) Use of the HCR-20 Version 3 with Women: A Narrative Synthesis, Journal of Forensic Psychology Research and Practice, 25:5, 930-959, DOI: 10.1080/24732850.2024.2386555

To link to this article: https://doi.org/10.1080/24732850.2024.2386555

Published online: 08 Aug 2024.

JOURNAL OF FORENSIC PSYCHOLOGY RESEARCH AND PRACTICE
2025, VOL. 25, NO. 5, 930–959
https://doi.org/10.1080/24732850.2024.2386555

Routledge
Taylor & Francis Group

Use of the HCR-20 Version 3 with Women: A Narrative Synthesis

Cevher Gunenc BSc, MSc, Lucy Fitton BSc, MSc, DClinPsy, and Peter Beazley BA, DClinPsy, MLaw

Norwich Medical School, University of East Anglia, Norwich, UK

ABSTRACT

The Historical-Clinical-Risk Management-20 Version 3 (HCR-20 V3) is a violence risk assessment tool that has been insufficiently studied in women. Existing reviews provide limited information on women's characteristics and clinical presentations. This review explores the profiles of women assessed with the HCR-20 V3. Five databases were searched and 11 studies included, totalling 445 women. The women mainly resided in Western inpatient psychiatric settings. Schizophrenia spectrum, substance misuse, and personality disorders were most common. Studies found the Clinical scale to be most predictive of future violence. Methodological limitations of the studies are discussed and recommendations made for future research and practice.

KEYWORDS

Risk assessment; risk factors; mental health; women; review

Introduction

Individuals detained in forensic and clinical settings, such as prisons or secure hospitals, are often done so for a range of purposes, with public safety and prevention of harm from violence being a major one (Yasrebi De Kom et al., 2022). Given its substantial estimated global cost and often devastating physical and psychological impact (Iqbal et al., 2021), clinicians, institutions, and policy makers have worked toward reducing and protecting society from the harm inflicted by violence, both within institutions and outside (Carpiniello et al., 2020). With this in mind, a range of structured professional judgment (SPJ) instruments, that combine both clinical judgment of professionals and risk assessment items derived from empirical research, have been developed to aid the prediction of violence and improve its accuracy relative to unstructured professional judgment and actuarial methods employed in the past (Douglas & Shaffer, 2020).

One such instrument developed for the assessment and management of violence risk is the Historical-Clinical-Risk Management-20 Version 3 (HCR-20 V3; Douglas et al., 2013). This is the most recent and updated version of the

CONTACT Peter Beazley p.beazley@uea.ac.uk Clinical Psychology Doctorate Programme, Department of Clinical Psychology & Psychological Therapies, Norwich Medical School, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich NR4 7TJ, UK

tool, comprising 10 Historical, five Clinical, and five Risk Management items found to be associated with future violence in those in forensic or mental health services (Guy et al., 2013). The Historical scale covers historical and static (or unchanging) items and the Clinical and Risk Management scales capture recent or future dynamic (or modifiable) items. Each item is then scored on a three-point scale (0 = not present/relevant, 1 = partially present/relevant, 2 = present/relevant), pertaining to whether or not the item is present for the individual and relevant to their risk of violence. Prior to the introduction of V3, items were only rated for presence, not relevance. A final summary risk judgment is made regarding the severity and imminence of future violence risk posed by the individual based on the information collected and reviewed.

Version 2 of the HCR-20 (V2; Webster et al., 1997) and its subscales have been repeatedly evaluated for predictive validity, which has since been established across a range of populations and settings (Guy & Wilson, 2007). V2 has become the most commonly implemented violence risk assessment tool across five continents (Singh et al., 2014). However, the release of V3 in 2013 introduced additional measures to manage risk as part of the assessment process, such as amendments to the items, relevance ratings of risk factors, risk formulation, scenario planning, and risk management planning (Douglas, 2014). It is likely that many institutions have begun to adopt the newest version of the HCR-20 (Logan, 2014), which makes it important to thoroughly research its utility and impact on risk assessment and management of individuals in both clinical and forensic settings.

Accurate and meaningful approaches to the identification of individuals who are most at risk of engaging in violence and those who would benefit most from intervention is of great consequence as it reveals targets and priorities for investment of finite resources and treatments to curb this adverse event (Whiting et al., 2021). This is in line with the Risk Need Responsivity (RNR) model (Andrews & Bonta, 2010) and SPJ approaches (de Vogel, De Beuf, et al., 2022), which stipulate that risk assessment should be individualized and enable risk management that is tailored to the unique needs of the person being assessed and treated, to prevent adverse behaviors. Available research suggests that men and women may have differing needs regarding risk assessment (de Vogel et al., 2019). One study found that women had employment and relationship differences to men, had more prior treatment for mental health conditions, had fewer convictions, and reported adverse childhood experiences more often than men (Streb et al., 2022).

Despite guidelines and potential gender differences, research with women has been neglected and understudied (Hughes, 2005). This is also true of other fields, for example in medicine where women are disadvantaged by the limited medical advancements and understanding of their specific health needs (Merone et al., 2022). However, it is arguably even more of an issue in the forensic context given that the large over-representation of males in the criminal justice system means

that research often excludes, or only partially represents, the female population (Clark, 2023). This has detrimental consequences in forensic and clinical domains, to both society, where female-perpetrated violence remains a risk, and to the women themselves. If their risk is overestimated, they may be detained under overly restrictive conditions that compromise their wellbeing and quality of life (Tully et al., 2023). If their risk is underestimated, they may be discharged too soon without adequate support, which could increase their vulnerability to adverse outcomes such as further offending, mental health difficulties or even mortality (Davies et al., 2007).

Women made up around 4% of those imprisoned in England and Wales in 2023, which was an increase from the previous year (Clark, 2023). Women in secure hospitals are as likely to perpetrate violence as men and engage in inpatient aggression more often than men (Jeandarme et al., 2017; Lieser & Rossdale, 2023). However, despite these findings and growing numbers of women in criminal justice systems internationally (Institute for Criminal Policy Research, 2022), most of what is known about the violence risk assessment and the utility of instruments such as the HCR-20 is informed by predominantly male samples (Gower et al., 2020). It is therefore unsurprising that studies employing a women-only sample assessed with the HCR-20 V2 typically report lower predictive validity (Ogonah et al., 2023).

As a solution to this, the Female Additional Manual (FAM; de Vogel et al., 2012) was developed to supplement the HCR-20 V2 when assessing women in forensic settings. It proposes that factors such as intimate partner violence, prostitution, pregnancy at a young age, parenting difficulties, and self-harm must be considered in addition to the 20 items assessed in men when predicting risk of violence in women. However, when the HCR-20 V2 has been supplemented with the FAM in women-only samples, the predictive validity remains lower than when the HCR-20 V3 is used alone (Ogonah et al., 2023). This suggests that the V3 amendments to the HCR-20 may have improved violence risk assessment for women, although this has not been replicated with larger or more diverse female samples.

Consistent with RNR and SPJ approaches, it is possible that the addition of risk formulation to the HCR-20 as well as assessing risk within contextual scenarios related to the individual's risk may have contributed to the improvement in risk prediction in women (de Vogel, De Beuf, et al., 2022). Risk formulation involves combining theory and empirically derived risk factors into a narrative that explains an individual's current difficulties in the context of their history and provides targets for modification of the factors that maintain these difficulties and risks (Tarpey et al., 2023). Especially in circumstances where individuals are nonresponsive to risk reduction interventions and their cases are deemed complex, formulation can support professionals in more holistic and person-centered decision-making regarding risk, restrictions, and treatment (Franke & Dudeck, 2019; Tarpey et al., 2023).

Risk assessment and formulation with the HCR-20 may achieve the desired outcome of violence reduction by informing treatment or risk management practices or by communicating complex and sensitive information to enable individuals and teams to co-formulate difficulties and challenges (The British Psychological Society, 2017). However, further information about the women requiring risk assessment and management interventions, their presentations, characteristics, and the settings and contexts that these needs arise in are required. In addition, it is inaccurate and ineffective to categorize all women in secure or mental health services within the same group. Understanding of women and their risk factors requires attention to the personal social and demographic characteristics that may intersect and compound risk, such as age, ethnicity, economic background, mental health needs, etc. (Montford & Hannah-Moffat, 2021).

Existing reviews have explored predictive validity of the HCR-20 V2 in women (Geraghty & Woodhams, 2015; Gower et al., 2020; O'Shea et al., 2013). Reviews that have expanded their inclusion to the predictive validity of the HCR-20 V3 either did not mention women in the samples (Challinor et al., 2021) or were only able to retrieve two studies of women assessed with V3, with little information about the characteristics of these women (Rossdale et al., 2020). Although it is essential to evaluate whether a tool can accurately predict the level of risk an individual poses and provide directions for reducing risk, clinicians must additionally be informed as to whether a tool is suitable for use with specific profiles, presentations, or settings. When selecting an instrument, clinicians should be supported by research that provides details about the individuals on whom studies assessing predictive validity have been conducted. Women's needs and circumstances may vary by ethnicity, age, or other factors, and the responses of clinicians and practitioners could be adapted accordingly to work with women in their care more effectively (Smith et al., 2020).

Despite the HCR-20 V3 being released 10 years ago, there are no reviews currently that summarize and describe the characteristics of women assessed with it, the types of studies that have investigated them, or the settings that this research took place in. Since the publication of Rossdale et al. (2020) review, which examined studies of women assessed with all versions of the HCR-20, it is possible that new papers may be available that could elucidate and characterize the women the HCR-20 V3 specifically has been implemented with.

The current review therefore aims to build an understanding of the profiles of women assessed with the HCR-20 V3 using reported individual characteristics and clinical presentations. It is anticipated that this will provide clinicians and practitioners with a relevant, recent, and clearer evidence base to make decisions about the generalizability of findings to the specific populations of women in their care, the suitability of using the HCR-20 V3 with these individuals, and the considerations that may be required when using the HCR-

20 V3 in practice. It is anticipated that this will subsequently improve the outcomes of women by enabling a more accurate and informative risk assessment, which increases the likelihood of improved, individualized risk management strategies that are least restrictive and most effective for the individual's goals, treatment, and care. A second objective of this paper is to build understanding of the types of studies, settings, and facilities that HCR-20 V3 is used in with women to enable clinicians to determine whether using the HCR-20 V3 is suitable within the specific setting in which they practice. Additionally, this paper aimed to also review and report findings of studies that have employed the HCR-20 V3 with women.

Methods

Design

A narrative synthesis design was selected for this review as it was best suited to the aims of exploring and describing the lesser investigated characteristics of women assessed with the HCR-20 V3. A narrative synthesis is a type of systematic review where databases are searched for relevant research and results of included studies are combined to describe, summarize, and explain the findings narratively (Popay et al., 2006). A narrative synthesis can be used when statistical synthesis is not possible due to limited available data (e.g. Campbell et al., 2018).

Registration

The current study utilized a systematic review and narrative synthesis design. Therefore, the review protocol was pre-registered in PROSPERO (CRD42023476561) and 2020 PRISMA guidelines (Page et al., 2021) were followed (Figure 1).

Search strategy

A search of the databases PsycInfo, MEDLINE Ultimate, CINAHL Ultimate, Scopus, and Web of Science was conducted on December 9, 2023 using the following search terms: ABS("Historic* Clinical Risk" OR "Historic* Clinical Risk Management" OR "HCR-20" OR "HCR-20V3") AND ABS("Version 3" OR "V3"). Reference lists of relevant reviews and papers were also scanned to identify additional studies, and full texts were retrieved of research published in 2013 or later, as this is when the HCR-20 V3 was released. Similar to Rossdale et al. (2020), search terms did not include "females" or "women" so as not to narrow the results or exclude mixed samples where women may have been included or discussed outside of the abstracts searched.

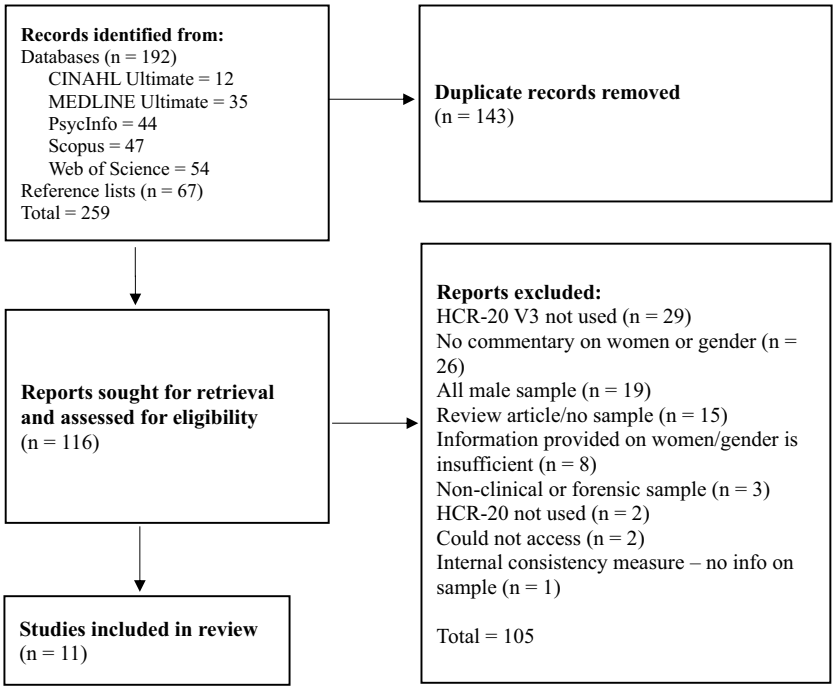

Figure 1. PRISMA flowchart (Page et al., 2021).

Study eligibility and selection

Studies were included if they were published research studies that have used the HCR-20 V3 in adult samples (aged 18 and over); their samples included women and these individuals were discussed separately to any men, or data were available for women separate to any men; they were conducted in a clinical or forensic setting, including facilities such as forensic hospitals, prisons, or community services; they were cross-sectional, longitudinal, cohort, case–control, retrospective, or prospective studies.

Studies were excluded if: they did not use the HCR-20 V3 to assess women; samples were of children or adolescents (aged 17 or under); samples were of only men; samples included a mix of men and women without commenting on the subsample of women or reporting any data on women separate to the men; they were review or opinion papers. As the HCR-20 V3 was released in 2013, relevant papers published prior to 2013 were still screened before being excluded.

Data extraction

Study characteristics (e.g. country, sample size, population and setting, type of study, how the HCR-20 V3 had been used and rated, outcome measures), sample characteristics (e.g. age, ethnicity, clinical presentations or mental

health conditions, previous offending and index offense, Mental Health Act or forensic status, other characteristics), and key findings (e.g. the profiles of women that the HCR-20 is being used with, the ability of the HCR-20 V3 in predicting violence and other outcomes in women compared to men, how the HCR-20 V3 performed compared to other instruments, the clinical utility of using the HCR-20 V3 with women in the sample) were extracted.

Where multiple time points were provided, mean age was extracted at index offense or admission rather than on assessment or discharge as this was the most frequently reported among the included studies. As different countries used different mental health legislation and criminal justice systems, the sections that individuals were detained under were considered forensic if they were court-ordered following an offense or the individuals were transferred to or from prison prior to or following hospital treatment. For studies that did not give the mean age of the subsample of women, this was estimated using the weighted average from the size of the sample of women and age ranges when these were provided. The most frequently reported effect size for predictive validity was Area Under the Curve (AUC), and this was extracted for each subscale of the HCR-20 V3, where provided. Where correlation coefficients were provided, an online effect size converter was used to estimate AUCs (Lin, 2020). One study (Campbell & Beech, 2018) reported negative binomial regression beta coefficients which were not possible to convert to AUC estimates. Another study reported chi-square coefficients which were also not able to be converted to AUCs (Crabtree, 2019).

Quality appraisal

The quality of studies was assessed using the JBI Checklist for Cohort Studies (Moola et al., 2020). The JBI Checklist evaluates the methodological quality of a study to determine the extent to which it has addressed the possibility of bias in its design, conduct, or analysis (Moola et al., 2020). Each study received a rating out of a maximum of seven points, as items 1, 2, 6, and 10 were deemed inapplicable to the included studies and were omitted. Due to high heterogeneity, the JBI was unable to appropriately capture the quality of some studies and one study that reported a single case was unable to be rated. However, other tools that were considered, including the Critical Appraisal Skills Programme (2023), Strengthening The Reporting of Observational Studies in Epidemiology (2024), and the Newcastle–Ottawa Quality Assessment Scale for cohort studies (NOQAS; Wells et al., 2014), had very similar items to the JBI and so the JBI was selected due to its brevity, clarity, and ability to derive a total rating for each study. All studies were rated independently by author C.G., and five were second-rated by author L.F. The authors addressed and resolved any uncertainties concerning the application of the JBI criteria to the studies included.

Results

Search results

Of the 116 records retrieved for review, 45 included women (Figure 1). Among these studies, there was a general trend of not reporting outcomes or conducting analyses of women's data separately to the men's. Some studies did not provide descriptive or demographic information for the women in their samples or explore or summarize any findings pertaining to similarities or differences between the men and women. Thirty-four studies that included women were therefore excluded, in addition to 71 other studies that did not meet the inclusion criteria. Eleven studies remained that met the inclusion criteria, totaling 445 women across the 11 samples. One study was a dissertation (Crabtree, 2019) and the remainder were peer-reviewed and published journal articles. The 11 studies were conducted across 8 different countries. Six of these studies included men and women and, although five of these contained limited information and conclusions regarding the women in the sample, they remained useful to include as study context, population, and setting could be extracted. Just under 60% of the total sample across the 11 included studies were women.

Quality appraisal

Quality appraisal ratings are provided in Table 1. One study was not able to be given a rating as it was a single-case study and the JBI items did not correspond well with its design. It was subsequently excluded from quality appraisal. Due to the variability of study designs, some leant themselves better to the JBI criteria than others. For example, two studies did not measure an outcome or conduct any analysis that the JBI asks about. For these studies, quality appraisal was still completed as interpretation could be made within the wider context of the overall results. Ten studies in total were given a rating between three and seven out of seven, with most of them receiving a rating of five. All studies that lost points did so due to not specifying whether confounding factors were identified, or which strategies were used to deal with them during analysis. The two studies that were given a rating of three out of seven also lost points due to not measuring or reporting follow-up. The JBI does not provide cutoffs regarding quality ratings (Moola et al., 2020) and no studies were excluded from this review based on their quality rating or inability of the tool to assess their quality.

Study characteristics

Table 1 presents features of each study (country, setting, type of study, outcomes, follow-up, instruments evaluated) and provides descriptive data about the studies (size, race or ethnicity, and mean age of the samples). Five studies did not provide information about the ethnicity of the sample or did not

Table 1. Characteristics of the 11 included studies.

Study, Country	Sample, n (% Women)	Mean age (range), Ethnicity	Type of study	Population, setting	Instrument(s) studied	How HCR-20 V3 was used, Rater	Quality rating
L. Campbell and Beech (2018), UK	89 (100)	35,[1] 51% Caucasian, 6% Black, 2% Asian, 1% Mixed	Prospective cohort study measuring predictive validity of HCR-20 V3	Psychiatric inpatients, Private secure forensic psychiatric hospital	HCR-20 V3, FAM	Routine clinical practice, Clinical team	5/7
Chen et al. (2023), China	152, 58 (38)	48 (25–75), 100% Chinese	Prospective cohort study comparing men and women, civil patients and offenders, and predictive validity of the HCR-20 V3 and VRS	Offenders with mental disorders and civil psychiatric patients, Prisons/ psychiatric hospitals	HCR-20 V3 (Chinese edition), VRS	Research purposes, Forensic psychology graduate research assistants	7/7
Cheng et al. (2023), Canada	32, 6 (19)	[1, 2, 3]	Retrospective cohort design evaluating how HCR-20 V3 items predict final risk formulations	Not criminally responsible (NCR) on account of mental disorder inpatients under warrant, Forensic psychiatric hospital	HCR-20 V3	Research purposes, Trained psychologists and senior undergraduate students	3/7
Chester et al. (2019), UK	84, 34 (41)	32 (18–57)[3]	Retrospective study examining the age at which antisocial or violent behavior was first exhibited	Inpatients with intellectual disability, Forensic intellectual disability inpatient service	HCR-20 V3	Routine clinical practice, Clinicians	3/7
Crabtree (2019), USA	18, 6 (33)	[1, 2, 3]	Longitudinal follow-up study comparing the predicted risk of recidivism between three tools	Clients diagnosed with a serious mental illness who have committed felony offenses, Community mental health clinic	HCR-20 V3, LSI-R, IORNS	Research purposes, Graduate student	5/7
de Vogel et al. (2019), The Netherlands	78 (100)	34 (20–65), 87% Caucasian	Retrospective follow-up study assessing predictive validity of HCR-20 V3	Forensic psychiatric inpatients, Forensic psychiatric hospitals	HCR-20 V3, HCR-20 V2, FAM, PCL-R, START, SAPROF	Research purposes, Trained and experienced psychologists and criminologists	5/7
Green et al. (2016), USA	124, 24 (19)	33[1], 21% Caucasian, 38% Black, 21% Hispanic, 17% "Other"	Retrospective file study comparing males and females	Forensic patients adjudicated Not Guilty by Reason of Mental Disease or Defect, State forensic hospital	HCR-20 V3	Research purposes, Trained psychology graduate research assistants	5/7
Lieser and Rossdale (2023), UK	42 (100)	36 (19-61), 57% Caucasian, 38% Black, 5% "Other"	Retrospective consecutive cohort design comparing HCR-20 V3 with FAM	Forensic psychiatric inpatients, Secure inpatient units	HCR-20 V3, FAM	Routine clinical practice, Qualified forensic psychologist	5/7

(*Continued*)

Table 1. (Continued).

Study, Country	Sample, n (% Women)	Mean age (range), Ethnicity	Type of study	Population, setting	Instrument(s) studied	How HCR-20 V3 was used, Rater	Quality rating
Mastromanno et al. (2018), Australia	40, 8 (20)	[1, 2, 3]	Retrospective follow-up of patient changes in HCR-20 V3 scores and relation with psychopathy and recidivism	Forensic psychiatric inpatients, Secure forensic mental healthcare facility	HCR-20 V3	Research purposes, Trained raters	7/7
Sorge et al. (2022), Italy	1 (100)	32, Italian	Single case study	Forensic psychiatric patient on probation, Therapeutic residential community	HCR-20 V3 (Italian edition), LSI	Research purposes, Patient's psychologist	N/A
Wolf et al. (2023), Germany	99 (100)	39 (18–67)[3]	Retrospective follow-up study assessing predictive validity of HCR-20 V3	Forensic psychiatric inpatients, Forensic psychiatric hospital	HCR-20 V3 (German edition), FAM	Research purposes, Trained raters	7/7

Note: HCR-20 = Historical-Clinical-Risk Management-20, FAM = Female Additional Manual, VRS = Violence Risk Scale, LSI-R = Level of Service Inventory-Revised, IORNS = Inventory of Offender Risk, Needs, and Strengths, PCL-R = Psychopathy Checklist – Revised, SAPROF = Structured Assessment of Protective Factors, START = Short-Term Assessment of Risk and Treatability, N/A = not applicable.
[1]Age ranges not provided or not provided separately for women
[2]Mean age not provided or not provided separately for women
[3]Ethnicity or race not provided or not provided separately for women

provide this separately for the subsample of women. Three studies did not provide the mean age for the women in their sample. The average ages of the remaining eight samples were observed to be older than 30 years. However, the age that was used by the eight studies differed between age at index offense (Green et al., 2016), age at admission (de Vogel et al., 2019; Sorge et al., 2022; Wolf et al., 2023), age at assessment (Campbell & Beech, 2018; Chen et al., 2023), or unspecified (Chester et al., 2019; Lieser & Rossdale, 2023).

Studies mainly utilized a retrospective design where case files were used to score the HCR-20 V3 for research purposes. Three studies did not investigate predictive validity or outcomes following assessment with the HCR-20 V3. The other eight studies primarily measured violence as an outcome, followed by recidivism and self-harm. These studies varied in the length of time they followed patients after assessment, ranging from 6 months to 12 years. This, as well as the type of violent outcome measured, may have resulted in the percentages of women who engaged in the outcomes differing substantially between studies, with rates of institutional violence, for example, ranging from 22% to 74%.

Mainly Western countries were represented in this review. The HCR-20 V3 was officially translated for use by researchers in China (Chen et al., 2023), Germany (Wolf et al., 2023), and Italy (Sorge et al., 2022) and then used to assess the individuals in each study. Four studies utilized the HCR-20 V3 with the FAM (Campbell & Beech, 2018; de Vogel et al., 2019; Lieser & Rossdale, 2023; Wolf et al., 2023), and two studies with the Level of Service Inventory (LSI; Crabtree, 2019; Sorge et al., 2022). Nine of the 11 studies used data from individuals residing in an inpatient forensic psychiatric setting.

Sample characteristics

It was anticipated that additional relevant sociodemographic information about the sample would also be available, such as financial, marital, family, education, employment, parenting, and accommodation status (Hamilton et al., 2017); however, these characteristics were collected/provided by only one study (Green et al., 2016). Further relevant characteristics that were unavailable from most of the published manuscripts included previous experiences of abuse or trauma, prior arrests or convictions, previous hospital admissions, and HCR-20 V3 individual item scores or data. Where data were combined with that of men in the sample, authors were contacted and separate demographic and HCR-20 V3 data for women were requested.

Percentages provided for the following sections reflect the proportion of women with each characteristic out of the total number of women in the samples that provided these data.

Race and ethnicity

Out of the five studies that provided the race or ethnicity of the women in their samples (Campbell & Beech, 2018; Chen et al., 2023; de Vogel et al., 2019; Green et al., 2016; Lieser & Rossdale, 2023), there appeared to be more women of White (61%), Black (19%), and Asian (35%) ethnic groups. This, however, does not provide a complete account of the numbers of women of each ethnicity across the total 445 women, due to this data being unavailable from six studies. All but one of the studies (Chen et al., 2023 was conducted in China) that provided information on ethnicity or race were conducted in Europe or the USA.

Diagnosis

Five studies provided the mental health condition or diagnosis given to the women in their samples (Campbell & Beech, 2018; de Vogel et al., 2019; Green et al., 2016; Lieser & Rossdale, 2023; Wolf et al., 2023), indicating that schizophrenia spectrum disorders (72%), substance misuse disorders (39%), and personality disorders (37%) were most prevalent. However, the women recruited by Wolf et al. (2023) all had a schizophrenia spectrum disorder, which may have skewed this finding. The least prevalent disorders across three studies (Campbell & Beech, 2018; Green et al., 2016; Wolf et al., 2023) were mood disorders (5%) and eating disorders (4%). Comorbidity was only measured by one study (Lieser & Rossdale, 2023).

Section and service

Four studies reported the type of section that patients were detained under (Campbell & Beech, 2018; de Vogel et al., 2019; Lieser & Rossdale, 2023; Wolf et al., 2023), with forensic sections (90%) being most common compared to civil sections (36%), across these four studies. Only two studies (Campbell & Beech, 2018; Chester et al., 2019), both conducted in the UK, reported whether the women in their samples resided on medium secure (37%), low secure (53%), or locked units (9%), as this is the security level system used across UK inpatient psychiatric hospitals.

Index offence

From the five studies that collected and reported the women's index offenses (de Vogel et al., 2019; Green et al., 2016; Lieser & Rossdale, 2023; Sorge et al., 2022; Wolf et al., 2023), violent offenses (43%) and homicide (36%) were the most common, followed by arson (17%), property (11%), and sexual offenses (2%) among the subsamples of women.

Predictive validity and outcomes

Predictive validity

Table 2 displays the reported AUC values of the Historical, Clinical, Risk Management, and Total HCR-20 V3 scales. Using the guidelines produced by Rice and Harris (2005) regarding classifications of effect sizes, the Total HCR-20 V3 AUC values from the included studies ranged from low (.635) to high (.840) effect size, indicating that a randomly selected female patient who has engaged in the outcome will score higher on the HCR-20 V3 than a randomly selected female patient who has not engaged in the outcome, in at least 63% of cases. This should be interpreted with caution, however, as these data have not been statistically combined and meta-analyzed within this review.

A particularly important observation is that most studies did not state whether they rated the R items in the context of the patient remaining within an institution or being discharged to the community. One rated the R items in the context of an institution (Campbell & Beech, 2018), two in the context of discharge to the community (Cheng et al., 2019; Crabtree, 2019), and one discharge to a facility of lower restriction (Green et al., 2016). Although some studies did rate and analyze relevance of items, studies mainly reported the presence of items. Summary risk ratings were not included by the majority of the studies and only three studies analyzed and reported them (Campbell & Beech, 2018; Chen et al., 2023; de Vogel et al., 2019).

Violence

Violence was the most common outcome measured among the included studies. Four out of the five studies that investigated violence as an outcome used the definition provided in the HCR-20 V3 manual (Douglas et al., 2013) to assess and define violence (de Vogel et al., 2019 used a definition of violent recidivism). However, the type of violence and when it was measured differed between institutional violence and violent recidivism on discharge. Nevertheless, Table 2 shows that the HCR-20 V3 Total scale scores were generally moderately accurate in discriminating between the women who engaged in violence from the women who did not within these studies. de Vogel et al. (2019) found that predictive validity for all types of recidivism was moderate, but low for violence. However, the HCR-20 V3 Total scale effect size for violent recidivism increased to a medium effect size in the 12-year follow-up period compared to the 3-year follow-up period, indicating that accuracy only increased over long periods of follow-up, which may not be useful to clinicians who may want to assess risk of more immediate violence. Lieser and Rossdale (2023) found larger effect sizes for the HCR-20 V3 subscales than the other studies and these were largest and significantly associated when predicting physical violence. In comparison, effect sizes for

Table 2. Predictive validity of the HCR-20 V3 for measured outcomes in women.

Study	*n*	Outcome, *n* who engaged in outcome (%)	Duration of follow-up	H	C	R	Total, Effect size classification
Chen et al. (2023)	58	Violence, 15 (26%)	6 weeks	.790	.770	.700	.850, High
			7–24 weeks	.770	.750	.660	.830, High
			6 months	.790	.730	.670	.840, High
de Vogel et al. (2019)	78	All recidivism (includes violent), 14 (18%)	3 years	.672	.680	.667	.711, High
		All recidivism (includes violent), 24 (31%)	12 years	.636	.673	.641	.667, Medium
		Violent recidivism, 6 (8%)	3 years	.604	.655	.544	.635, Low
		Violent recidivism, 13 (17%)	12 years	.649	.690	.618	.672, Medium
Green et al. (2016)	24	Institutional violence, 13 (54%)	14 months	.660	.678	.455	.654, Medium
Lieser and Rossdale (2023)	42	Physical violence, 31 (74%)	6 months	.713	.739	.730	.702, Medium
		Non-physical violence, 11 (26%)		.581	.597	.516	.710, High
		Any violence, 42 (100%)		.610	.661	.541	.714, High
Wolf et al. (2023)	99	Violent index offense, 80 (81%)	10 years	Subscale AUCs not provided			AUCs not provided
		Institutional violence, 22 (22%)					AUCs not provided
		Violent recidivism, 9 (10%)					.695, Medium

Note: n = number of women, H = Historical subscale, C = Clinical subscale, R = Risk Management subscale, AUC = area Under the curve

nonphysical violence, defined as verbal abuse and threats, were smaller and were not significantly associated. The Total scale predictive validity was largest for any violence (combined physical and nonphysical violence). In three studies that provided subscale AUCs (de Vogel et al., 2019; Green et al., 2016; Lieser & Rossdale, 2023), the predictive validity of the Total scale was outperformed by the Clinical scale, which yielded the highest AUC values in comparison to the other scales. This was in contrast to the findings by Chen et al. (2023) who reported larger effect sizes in the Total and Historical scales. The Risk Management scale, across all four studies, however, demonstrated lower predictive validity compared to the other scales.

Two studies analyzed predictive validity of individual items of the HCR-20 V3 in women. Measuring the outcome of institutional violence, Green et al. (2016) reported that the Historical item *history of problems with other antisocial behaviour* and the Clinical item *recent problems with instability* were significantly associated with violence in women. Similarly studying the outcome of institutional violence, Wolf et al. (2023) found that the Clinical items *recent problems with violent ideation or intent, recent problems with instability*, and *recent problems with treatment or supervision response* were significantly associated with institutional violence. Violent recidivism was significantly associated with the Historical item *history of problems with other antisocial behaviour – during childhood*, the Clinical items *recent problems with insight – treatment need* and *recent problems with instability – cognitive*, and the Risk Management item *future problems with treatment or supervision response.*

Three studies compared the HCR-20 V3 used alone, with the HCR-20 V3 used with the FAM in predicting violence in women-only samples (de Vogel et al., 2019; Lieser & Rossdale, 2023; Wolf et al., 2023). The HCR-20 V3 was found to have larger predictive validity alone than when used with the FAM in all three of the studies, indicating that the FAM was not as able to discriminate between violent and nonviolent women. When predicting physical violence without the FAM, Lieser and Rossdale (2023) reported larger effect sizes for each subscale and the Total scale of the HCR-20 V3, whereas the FAM was found to reduce predictive validity when used alongside each subscale and Total scale of the HCR-20 V3. Predictive validity did not differ between the HCR-20 V3 and the FAM when predicting any and nonphysical violence. Two studies measured the LSI with the HCR-20 V3; however, either scores for women were not provided separately (Crabtree, 2019), or were provided without comparative analysis to the HCR-20 V3 (Sorge et al., 2022). Interestingly, the Short-Term Assessment of Risk and Treatability (START) Vulnerability scale, measured only by de Vogel et al. (2019), predicted violent recidivism with larger effect size in both the three-year follow-up period (.697), and in the 12-year follow-up period (.704), when compared to the HCR-20 V3. The HCR-20 V3 remained performing better than the HCR-20 V2 at both the 3-year follow-up period (.563) and the 12-year period (.592), in this sample of women, however (de Vogel et al., 2019).

Four studies compared differences in the HCR-20 V3 as a function of gender. Green et al. (2016) found that a higher proportion of women in the sample engaged in any and physical violence compared with men in the sample; however, this was not significantly different. The authors also reported a stronger relationship between scale scores and violence among men than women, although the difference was only significant for the Historical item *violent attitudes*, which was only significantly related to violence in men. Women in the sample were less likely to have been previously arrested and more likely to have been married previously compared with the men, although they scored significantly higher than men on the items *history of problems with relationships* and *history of problems with traumatic experiences*. Women were more likely to have been diagnosed with a mood disorder, although the most common diagnosis among the women was borderline personality disorder. Although no differences were observed between genders in associations between HCR-20 V3 scores and violence, authors reported that interrater agreement tended to be higher when rating men compared to women. Chen et al. (2023) reported that the effect sizes in women were significantly smaller on the Risk Management scale (relevance scores) compared to effect sizes in men. However, on the Clinical (presence and relevance scores) and Total scales (presence scores), effect sizes were significantly larger in women than men. The authors also found that the rate of violence was higher in women than men; however, as the follow-up period increased, this reversed, and men were found to engage in more violence than women. Effect sizes remained higher for women across all time periods when presence ratings were used and were similar between men and women when relevance ratings were used. Crabtree (2019) found that none of the women in the sample, compared with 17% of the men were classified as high risk on the HCR-20 V3. They found that three women were classified as low risk and the other three as moderate risk. Mastromanno et al. (2018) found that women's mean scores were lower than men's on the Clinical and Risk Management scales of the HCR-20 V3, both at the beginning and end of treatment, although this difference was not statistically significant.

Other outcomes

L. Campbell and Beech (2018) was the only study exploring the predictive validity of the HCR-20 V3 for self-harm in women. The Historical and Clinical scales were significantly positively associated with the frequency of self-harm, although the Risk Management scale was not. The HCR-20 V3 Total score was significantly positively associated with self-harm, with or without the FAM; however, the association was slightly stronger (nonsignificant) with the FAM items included.

Although mortality was not planned as an outcome measure at the outset, it was analyzed post-hoc by de Vogel et al. (2019) following a high rate of death among the women in the sample during the follow-up period ($n = 14$, 18%). Mean age of death among the women who died was 45. The HCR-20 V3 was a weak predictor of mortality (AUC = .607), with the Historical scale resulting in the largest effect size (.605) compared to the Clinical (.585) and Risk Management (.600) scales. The FAM alone and with the HCR-20 V3 yielded low effect sizes (.550 and .608, respectively). Notably, the Psychopathy Checklist-Revised (PCL-R) Interpersonal facet was a significant protective factor for mortality (.308).

Three studies explored nonviolent recidivism as an outcome. Unfortunately, separate data for women were not provided by Crabtree (2019) or Mastromanno et al. (2018) and the conclusions drawn from the analyses are likely to have better represented the subsamples of men as the subsamples of women in both studies were much smaller (33% and 20%, respectively). de Vogel et al. (2019) evaluated the predictive validity of the HCR-20 V3 compared with the other tools such as the FAM in an all-women sample. In the 3-year follow-up period, authors reported a higher AUC effect size for all recidivism (violent and nonviolent) when the HCR-20 V3 was used alone (.711) compared to when the FAM was used alone (.676) or the two tools were used together (.695), although all three were significantly predictive of all recidivism. Interestingly, when the FAM final risk judgments for nonviolent criminal behavior were used alone to predict nonviolent recidivism, effect sizes increased to .860 over the 3-year period.

Discussion

Women comprise an increasing subgroup of clinical and forensic service users but there is limited research on the use of SPJ risk assessment tools, such as the HCR-20 V3, with women as these tools have mainly been evaluated with men. This systematic review and narrative synthesis aimed to investigate the characteristics of the women assessed with the HCR-20 V3 and the contexts in which these studies occurred. Eleven papers included women in their samples who had been assessed with the HCR-20 V3, enabling preliminary exploration of their profiles and the clinical and forensic settings they occupy. However, many of these studies used a relatively small sample of women, limiting external validity and generalizability (Steyerberg et al., 2003). In addition, there was great variability among the studies in the characteristics each provided, the outcomes measured, and the durations of follow-up, rendering comparison and summary of studies difficult. Areas of similarity across the studies included use of inpatient forensic psychiatric samples, retrospective cohort designs, and the violence definition among the studies exploring outcomes.

Study and sample characteristics

Studies were mainly conducted in Europe and North America, which is consistent with findings of other reviews that studied risk assessment of women (Geraghty & Woodhams, 2015; Rossdale et al., 2020). The finding that most women were of White, Black, or Asian ethnicity, the latter of which mainly comprised Chen et al. (2023) sample, may reflect the geographical locations of these studies. However, six studies did not provide race or ethnicity information, amounting to a third of the sample of women. This highlights the ongoing need for further research and improvement in reporting of characteristics such as this, to explore whether findings can be generalized across different countries and ethnicities. In addition, there is a need to ensure that risk assessments, found to be moderately racially biased in men (Monjazeb & Douglas, 2022), do not disadvantage women of ethnic minority groups by biasing them toward higher risk classifications.

Papers largely represented an inpatient forensic psychiatric population, with only two studies being conducted in a community setting and no female samples were from a prison setting, potentially reflecting the higher numbers of women in psychiatric hospitals versus prisons and the need for violence risk assessment in inpatient contexts compared with community services. This is consistent with Rossdale et al. (2020) review, which included fewer studies in civil psychiatric and prison settings, but contrasts with Geraghty and Woodhams (2015) and Gower et al. (2020) who observed that the majority of the samples in their reviews were correctional. This may reflect differing inclusion criteria.

Within the studies that provided this information, the majority of women were detained under forensic sections and resided on low secure units, suggesting a history of previous offending and violence, with a level of risk potentially assessed as low enough to not necessitate conditions of medium security. This is supported by a violent index offense being committed by around half of the women in the studies that reported this. During the follow-up period, the reported number of women who engaged in physical violence was 94 (31%), indicating that a minority of women in the samples continued to engage in violence. Similar findings were observed in previous reviews, which also noted that less than half of the females across their samples engaged in violence over the follow-up period (Geraghty & Woodhams, 2015; Gower et al., 2020).

The mean ages of the women in this review ranged from 32 to 48 across the studies that provided this, which was somewhat consistent with other reviews (Geraghty & Woodhams, 2015; Rossdale et al., 2020), though both the lower and upper mean ages in the previous reviews were younger than in the current review. It was not possible to compare age of the samples of women with the

samples of men in the current review as five studies did not include men and four studies did not provide separate mean ages for the men. However, previous research comparing men and women have found that women at admission or court verdict are generally older than men in some samples, though not significantly so (de Vogel, Stam, et al., 2022; Dean et al., 2020). The older age detected in this review may be due to women being in institutions for some time before they were assessed with the HCR-20 V3, especially as this version was not released until 2013 and most studies used a retrospective design.

A retrospective design examining file information was primarily utilized among the studies, with only two papers using a prospective cohort design. A retrospective design enables data collection, analysis, and evaluation in a shorter timeframe without waiting for the follow-up period to elapse, thus reducing project duration and costs (Talari & Goyal, 2020). However, it increases the risk of missing data, poorer quality data, and participant attrition (Talari & Goyal, 2020). The authors of the HCR-20 V3 recommend the use of prospective study designs when assessing the predictive validity of the HCR-20 V3 (Douglas et al., 2013). Additionally, retrospective coding of items using file information conflicts with the intended use of SPJ tools such as the HCR-20 V3, which advocate incorporating professional judgment by the clinical team based on individualized and collaborative formulation and knowledge of the person beyond written reports (de Vogel, De Beuf, et al., 2022). Therefore, risk derived from totaling scores rated by researchers who likely never interacted with the individual aligns more with actuarial risk assessment methods that lack individualization and targets for risk management (Douglas, 2014).

This is of particular importance for women, who are found to have more complex pathways to recidivism, such as victimization, social marginalization, poorer relational functioning, and mental health problems (Brennan et al., 2012). For example, women are diagnosed with borderline personality disorder at a higher rate than men in the clinical population (Chapman et al., 2022; Huang et al., 2009) and were diagnosed with borderline personality disorder in over a third of the cases across five studies in this review. Risk assessment of women in general, especially those with borderline personality disorder, requires attention to subtler relational and interpersonal patterns that may have direct bearing on risk and management practices (Bohus et al., 2021; Whiting et al., 2021), which can be easily overlooked in written reports. Relying on file information without patient interview and team formulation, contrary to the HCR-20 V3 process (Douglas et al., 2013), may compromise risk assessment accuracy and subsequent management strategies (Hopton et al., 2018) and is not generalizable to contexts where the HCR-20 V3 may be used as recommended in routine clinical practice.

The remaining women across the five studies had a diagnosis of mainly schizophrenia spectrum disorders and substance misuse. It is unclear whether

this was also the case in previous reviews as clinical or psychiatric populations of women were not the focus of these and frequencies of each diagnosis were not provided. However, in a national survey of mental health in England, 14% of people screened positive for any personality disorder, whereas less than 1% of people had a psychosis-related disorder (NHS, 2018). Wolf et al. (2023), whose paper was included in the current review, investigated a forensic psychiatric sample of women with a primary schizophrenia spectrum diagnosis, which could have inflated the total. The authors reported that around half of the sample also had a diagnosis of personality disorder or substance misuse. Any comorbidity was unspecified by three of the five other studies in this review; however, dual diagnosis is a common finding across psychiatric inpatient samples and there is often an overlap of schizophrenia with substance misuse and personality disorder (Howner et al., 2018).

Using the HCR-20 V3 with women

Due to high variability in the designs and methods of the included studies, and a general tendency in research to analyze data from men and women together to maximize statistical power, the use of the HCR-20 V3 with women was difficult to synthesize as limited information and conclusions could be extracted about women specifically that were similar across multiple studies.

Although many of the follow-up studies measured violent outcomes, variability in measures and procedures may have resulted in predictive validity of the HCR-20 V3 for violence in women also differing vastly between the studies. Effect sizes ranged from low to high for the total scale of the HCR-20 V3. The studies that found high effect sizes at six-month follow-up hold promise for accurate risk assessment of women in clinical and forensic services; however, only two studies found this, and both used a relatively small sample of women. Studies that found lower effect on violence followed women for longer durations, which may explain the reduced accuracy as the HCR-20 is intended to assess risk of future violence over a six-month period (Douglas et al., 2013). Studies that have evaluated the HCR-20 V3 with men have produced more consistent findings of larger effect sizes for the prediction of violence (Smith et al., 2020), indicating that the reliability of findings remain reduced for women and require further replication and with larger samples to improve this.

When attending to individual scales, studies reported that the Clinical scale was generally the most accurate predictor of violence in women, while the Risk Management scale was the least. This contrasts with studies of predominantly male samples that report predictive validity and larger effect sizes across all subscales (Doyle et al., 2014). However, many of the included studies rated Risk Management items, intended as predictors of future risk, in retrospect, which may have been difficult to

assess using historic file information, thus reducing the predictive validity of this scale. In addition, many of the studies did not specify whether the items were rated in the context of individuals remaining in the institution or being discharged to the community in the next 6 months. This may have affected the performance of the scale if assessment and consideration of factors did not correspond with the future treatment or discharge plans for the individual. It therefore remains unclear without further research that addresses these limitations to ascertain whether the Risk Management scale adds predictive value to the risk assessment of violence in women. However, it is likely that current and modifiable risk factors, such as those assessed by the Clinical scale, are particularly useful and relevant to assess in women.

Only two studies explored individual items, reporting that previous anti-social behaviors and recent instability were most associated with violence. Both factors are consistent with previous research on lifestyle and personality features that may make violence more likely in women's lives. Poor emotional and behavioral control and diverse criminal offending may put women in risky situations that result in both violent victimization and perpetration (Turanovic et al., 2015). This is worth bearing in mind as interventions provided within an institution may help to improve instability while the individual is residing there, whereas lifestyle factors and criminogenic social environments are much more difficult to improve from within an institution. These vulnerability factors for future violence may require community service in-reach and holistic social care packages that include financial, housing, and mental health support to maintain desistance (Gålnander, 2020). This is especially true for older women and those of ethnic minority or nonwhite groups, of which the included samples comprised, who may have lived with social exclusion and disadvantage for longer periods than their counterparts, especially in Western countries. Being a woman with these sociodemographic characteristics produces cumulative disadvantage (Mann, 1989). This may result in a weakened social position in society, which increases barriers to desistance and social inclusion (Bersani & Doherty, 2018). Barriers include limited work and educational opportunities, poorer health outcomes, smaller social network, and poverty (Hinze et al., 2012), that wider policy and social care support could help with.

The risk assessment tool that comes closest to capturing risk factors specific to women is the FAM, albeit without direct consideration of these particularly disadvantaged subgroups, although some of the items may reflect the effects of lived disadvantage, e.g. low self-esteem, suicidality/self-harm. The tool makes efforts to consider issues found to be particularly relevant to the risk behavior of women in general within clinical and forensic settings. However, this review found that the FAM, when used by itself, did not perform as well in predicting violence as the HCR-20 V3

alone. In fact, when used in conjunction with the HCR-20 V3, many studies found that the FAM reduced the significance and effect size, especially for outcomes of physical violence. This led to the studies concluding that, although the FAM may have been a valuable addition to the HCR-20 V2, the amendments to the HCR-20 V3 may have slightly improved the accuracy of risk assessment with women (de Vogel et al., 2019), although the mechanisms of this improvement remain unclear without further exploration of the assessment process with the HCR-20 V3 and the specific aspects that add value and accuracy to the prediction of violence risk in women.

Limitations of the literature

The exploration of existing literature revealed several limitations that impact the generalizability and robustness of findings regarding the use of the HCR-20 V3 with women. Despite the increasing numbers of women in clinical and forensic settings, there remains a scarceness of research focused on the use of risk assessment tools with women specifically. Many studies included in this review featured relatively small samples of women using heterogenous methods, which limit the external validity and generalizability of results, and reduce the clarity of any conclusions. The majority of studies were conducted in Europe and North America, reflecting a narrow geographical focus that may also limit the generalizability of findings to other regions, nationalities, and cultures. Additionally, a substantial proportion of studies did not provide information on ethnicity, impacting the ability to consider the use of the HCR-20 V3 with a diverse female population.

Quality appraisal of the studies revealed that many did not measure or control for psychosocial or demographic factors such as age, ethnicity, diagnosis, comorbidity, financial status, marital status, education, employment, housing status, presence of substance addiction, or presence of trauma, previously found to be relevant or potentially confounding (Hamilton et al., 2017). This would have enabled evaluation of whether and how predictive validity differed as a function of these factors and better ascertained the capacity of the HCR-20 V3 in assessing risk in all women or select subgroups. This is especially important as some groups of women were underrepresented in the included samples, such as women of ethnic minority backgrounds, women in prison, community mental health or community forensic settings, and women who had a diagnosis other than schizophrenia spectrum, substance misuse, or personality disorders.

In addition to issues of representativeness and generalizability, the way in which the HCR-20 V3 risk assessments were conducted in some studies highlighted the potential absence of implementation of the tool's newest version updates. An important addition to the HCR-20 V3 was the relevance

ratings for each item that allow for assessment of not only presence but also of whether a risk factor directly contributed to past violence, has any bearing on decision-making that promotes violence, and is critical to risk management plans (Douglas, 2014). Although some studies utilized relevance ratings, most only reported and analyzed the presence of items. This indicates that the full intended features of the HCR-20 V3 are not being applied by researchers or clinicians. Moreover, Chen et al. (2023) found that effect sizes differed less between men and women when relevance ratings were used, suggesting that using relevance ratings could improve the accuracy of risk assessment for women, making it potentially comparable to the assessment accuracy observed for men. The use of relevance ratings thus remains an important are for future research to consider.

An additional shortcoming of some studies was their neglect to assess or report summary risk ratings, which allow the rater to make an overall judgment of the level of risk of the individual after careful consideration of all factors (Douglas et al., 2013). Prior research has shown that the summary risk ratings used by SPJ instruments are as, or more, accurate than the use of actuarial instruments and have been found to improve predictive validity in male samples (Douglas, 2014). Thus, by overlooking them in assessment and analysis, predictive validity of the HCR-20 V3 in women may have been underestimated.

A further issue with the use of the HCR-20 V3 in the included studies is the impact of not specifying whether the Risk Management scale was scored in the context of discharge to the community in the next 6 months or remaining within an institution. As has been discussed, this may have affected the low effect sizes found for this scale, which may have resulted from being unable to accurately assess whether the five risk factors in the scale will be present or relevant to the individual in the upcoming months. For example, the score for the items *future problems with living situation* and *future problems with personal support* may very well increase if rated prior to community discharge where an individual has previously found it difficult to remain in stable and suitable accommodation or plans to move to an area that is a considerable distance from protective family members or friends. These items would have been especially difficult to rate accurately in retrospect using file information, which was the design of many of the included studies.

Conclusions

This review aimed to shed light on the characteristics of women and the existing literature that has investigated the application of the HCR-20 V3 in assessing women's risk. There was a small number of eligible studies, and they utilized relatively small samples of women. Within this limited scope, however, this review found that most research on the HCR-20 V3 used with

women took place in inpatient forensic psychiatric settings where the majority of women were detained under forensic sections and resided on low secure units. The mean ages of the women ranged from 32 to 48 and most women had a schizophrenia spectrum disorder, substance misuse diagnosis, or a personality disorder.

Where studies evaluated predictive validity, effect sizes for violence in women were lower than effect sizes reported in previous research with men. However, effect sizes were not pooled and meta-analyzed due to insufficient data and research available currently and therefore these findings could be the result of methodological issues with the implementation of the HCR-20 V3 in a research context. Without further data and analysis, however, it remains unclear whether these findings are due to these methodological limitations or reflect the inability of the HCR-20 V3 in predicting violence as strongly and reliably in women as it does in men. Therefore, until further research on the predictive validity of the HCR-20 V3 in women is conducted, it is recommended that clinicians continue with individual formulation and creating risk management and care plans informed by formulation and treatment needs identified from both clinical assessment and risk assessment with the HCR-20 V3 that utilizes the updated features as intended for use.

Analysis of individual items indicate that emotional, behavioral, or cognitive instability are particularly relevant in women's future risk of violence and may highlight areas for interventions to target, e.g. emotional and behavioral regulation strategies. However, these may only go so far as to address institutional risk as violence in the community is likely to be precipitated by lifestyle and psychosocial factors.

The methodological designs of the included studies may explain some of these findings, which may not have captured the full spectrum of characteristics of women, the settings they occupy, and the full utility of risk assessment with the HCR-20 V3. Nonetheless, this review supports that gender differences may exist with regard to violence and offending and further research on women's risk and assessment is warranted.

Recommendations for research and practice

Limitations identified in the existing literature highlight the need for further research that addresses these shortcomings. Future studies should aim to study larger and more diverse female samples, keep any women in the sample without excluding them, and report and describe their characteristics, even if analysis is not possible. Where separate analysis of women's data is possible, merely including gender as a control variable in statistical models is inadequate and could lead to potentially important differences between men and women being missed (Turanovic et al., 2015). Future studies should also fully incorporate the updates of the

HCR-20 V3 by using and evaluating relevance ratings, summary risk ratings, risk formulations, and scenarios that have not yet been sufficiently investigated. Predictive validity of these processes of the HCR-20 V3 has been analyzed by some studies (Chen et al., 2023 have analyzed relevance ratings and Guy (2008) have analyzed summary risk ratings for other SPJ tools). Researchers can review the methodology of these studies as well as the User Guide of the HCR-20 V3 (Douglas et al., 2013) for further information on conducting predictive validity evaluations of these aspects of the HCR-20 V3. Ideally, studies would be conducted prospectively, utilizing HCR-20 V3 assessments that have been completed in routine practice as opposed to retrospectively using file information without interview of the individual. Studies should consider and specify in their manuscripts the context in which Risk Management items were assessed. Studies need to record and report comprehensive sociodemographic information and identify subgroups of women to explore the effects of intersecting characteristics on predictive validity.

Clinically, as noted in the relevant manual, professionals should rate items based on both presence and relevance, the latter of which may assist in making judgments of overall risk. Clinical use of the relevance ratings may allow further retrospective research to better establish the importance of these ratings. Individuals should be interviewed to inform their risk assessment and allow the opportunity for collaborative formulation and treatment planning. Individualized formulations should be developed and used to generate solutions and plans to manage and prevent risk. The findings from this review do not support the use of the FAM; however, professionals may wish to assess and be aware of these additional factors that may impact risk and be of relevance to individual formulations.

Disclosure statement

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

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

* Indicates the references of the studies that were included in this review

Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, & Law*, *16*(1), 39. https://doi.org/10.1037/a0018362

Bersani, B. E., & Doherty, E. E. (2018). Desistance from offending in the twenty-first century. *Annual Review of Criminology*, *1*(1), 311–334. https://doi.org/10.1146/annurev-criminol-032317-092112

Bohus, M., Stoffers-Winterling, J., Sharp, C., Krause-Utz, A., Schmahl, C., & Lieb, K. (2021). Borderline personality disorder. *Lancet*, *398*(10310), 1528–1540. https://doi.org/10.1016/S0140-6736(21)00476-1

Brennan, T., Breitenbach, M., Dieterich, W., Salisbury, E. J., & Van Voorhis, P. (2012). Women's pathways to serious and habitual crime: A person-centered analysis incorporating gender responsive factors. *Criminal Justice & Behavior*, *39*(11), 1481–1508. https://doi.org/10.1177/0093854812456777

The British Psychological Society. (2017). *Practice guidelines* (3rd ed.). https://www.bps.org.uk/guideline/bps-practice-guidelines-2017

* Campbell, L., & Beech, A. (2018). Do scores on the HCR-20 and FAM predict frequency of self-harm in females within a secure psychiatric hospital? *Journal of Forensic Psychiatry & Psychology*, *29*(6), 914–933. https://doi.org/10.1080/14789949.2018.1477975

Campbell, M., Katikireddi, S. V., Sowden, A., McKenzie, J. E., & Thomson, H. (2018). Improving conduct and reporting of narrative synthesis of quantitative data (ICONS-Quant): Protocol for a mixed methods study to develop a reporting guideline. *British Medical Journal Open*, *8*(2), e020064. https://doi.org/10.1136/bmjopen-2017-020064

Carpiniello, B., Mencacci, C., & Vita, A. (2020). Violence as a social, clinical, and forensic problem. *Violence and Mental Disorders*, 3–24. https://doi.org/10.1007/978-3-030-33188-7

Challinor, A., Ogundalu, A., McIntyre, J. C., Bramwell, V., & Nathan, R. (2021). The empirical evidence base for the use of the HCR-20: A narrative review of study designs and transferability of results to clinical practice. *International Journal of Law and Psychiatry*, *78*, 101729. https://doi.org/10.1016/j.ijlp.2021.101729

Chapman, J., Jamil, R. T., & Fleisher, C. (2022). *Borderline personality disorder. National Library of Medicine; StatPearls Publishing*. https://www.ncbi.nlm.nih.gov/books/NBK430883/

* Chen, Y., Douglas, K. S., Zhang, Z., Xiao, C., Wang, H., Wang, Y., & Ma, A. (2023). Evaluating the HCR-20V3 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

* Cheng, J., Haag, A. M., & Olver, M. E. (2019). Predictors of historical clinical risk management-20 version 3 (HCR–20: V3) summary risk ratings. *Psychiatry, Psychology and Law*, *26*(4), 682–692. https://doi.org/10.1080/13218719.2019.1618753

* Chester, V., Wells, H., Lovell, M., Melvin, C., & Tromans, S. J. (2019). The prevention of offending behaviour by people with intellectual disabilities: A case for specialist childhood and adolescent early intervention. *Advances in Mental Health and Intellectual Disabilities*, *13*(5), 216–227. https://doi.org/10.1108/AMHID-03-2019-0008

Clark, D. (2023). *Number of prisoners in custody in England and wales from 1900 to 2023, by gender*. Statista. https://shorturl.at/bkvFT

* Crabtree, T. S. (2019). *Risk assessment tools and mental health courts: Preliminary findings* [Unpublished doctoral dissertation]. Spalding University. https://shorturl.at/yHSUZ.

Critical Appraisal Skills Programme. (2023). *CASP cohort study checklist*. https://casp-uk.net/checklists/casp-cohort-studies-checklist-fillable.pdf

Davies, S., Clarke, M., Hollin, C., & Duggan, C. (2007). Long-term outcomes after discharge from medium secure care: A cause for concern. *British Journal of Psychiatry*, *191*(1), 70–74. https://doi.org/10.1192/bjp.bp.106.029215

Dean, K., Singh, S., Kemp, R., Johnson, A., & Nielssen, O. (2020). Characteristics and re-offending rates amongst individuals found not guilty by reason of mental illness

(NGMI): A comparison of men and women in a 25-year Australian cohort. *International Journal of Forensic Mental Health*, *20*(1), 17–30. https://doi.org/10.1080/14999013.2020.1795011

* de Vogel, V., Bruggeman, M., & Lancel, M. (2019). Gender-sensitive violence risk assessment: Predictive validity of six tools in female forensic psychiatric patients. *Criminal Justice & Behavior*, *46*(4), 528–549. https://doi.org/10.1177/0093854818824135

de Vogel, V., De Beuf, T., Shepherd, S., & Schneider, R. D. (2022). Violence risk assessment with the HCR-20V3 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., de Vries Robbé, M., van Kalmthout, W., & Place, C. (2012). Risk assessment of violent women: Development of the 'female additional manual' (FAM). *Tijdschrift voor psychiatrie*, *54*(4), 329–338. https://shorturl.at/hmAI8

de Vogel, V., Stam, J., Bouman, Y. H. A., Ter Horst, P., & Lancel, M. (2022). Gender differences in substance abuse history and offending behavior: A multicentre study in Dutch forensic psychiatry. *Journal of Forensic Psychology Research and Practice*, *22*(1), 1–17. https://doi.org/10.1080/24732850.2021.1945833

Douglas, K. S. (2014). Version 3 of the historical-clinical-risk management-20 (HCR-20V3): Relevance to violence risk assessment and management in forensic conditional release contexts. *Behavioral Sciences & the Law*, *32*(5), 557–576. https://doi.org/10.1002/bsl.2134

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

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 K. S. Douglas, & R. K. Otto (Eds.), *Handbook of Violence Risk Assessment* (2nd ed., pp. 253–293). Routledge.

Doyle, M., Power, L. A., Coid, J., Kallis, C., Ullrich, S., & Shaw, J. (2014). Predicting post-discharge community violence in England and wales using the HCR-20V3. *International Journal of Forensic Mental Health*, *13*(2), 140–147. https://doi.org/10.1080/14999013.2014.906517

Franke, I., & Dudeck, M. (2019). What might work when nothing seems to work: Case formulation in the treatment of antisocial personality disorder in a forensic mental health setting. In *Case formulation for personality disorders* (pp. 161–180). Academic Press. https://doi.org/10.1016/B978-0-12-813521-1.00009-6

Gålnander, R. (2020). Desistance from crime—to what? Exploring future aspirations and their implications for processes of desistance. *Feminist Criminology*, *15*(3), 255–277. https://doi.org/10.1177/1557085119879236

Geraghty, K. A., & Woodhams, J. (2015). The predictive validity of risk assessment tools for female offenders: A systematic review. *Aggression & Violent Behavior*, *21*, 25–38. https://doi.org/10.1016/j.avb.2015.01.002

Gower, M., Spiranovic, C., Morgan, F., & Saunders, J. (2020). The predictive validity of risk assessment tools used in Australia for female offenders: A systematic review. *Aggression & Violent Behavior*, *53*, 101428. https://doi.org/10.1016/j.avb.2020.101428

* Green, D., Schneider, M., Griswold, H., Belfi, B., Herrera, M., & DeBlasi, A. (2016). A comparison of the HCR-20V3 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.

Guy, L. S. (2008). *Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey* [Unpublished doctoral dissertation]. Simon Fraser University.

Guy, L. S., & Wilson, C. M. (2007). *Empirical support for the HCR-20: A critical analysis of the violence literature. HCR-20 violence risk assessment white paper series, no. 2*. Mental Health, Law, and Policy Institute, Simon Fraser University.

Guy, L. S., Wilson, C. M., Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). HCR-20 V3: Item-by-item summary of violence literature. In *HCR-20 violence risk assessment white paper series* no 3. Mental Health. Law, and Policy Institute, Simon Fraser University.

Hamilton, Z., Campagna, M., Tollefsbol, E., Van Wormer, J., & Barnoski, R. (2017). A more consistent application of the RNR model: The STRONG-R needs assessment. *Criminal Justice & Behavior*, *44*(2), 261–292. https://doi.org/10.1177/0093854816678032

Hinze, S. W., Lin, J., & Andersson, T. E. (2012). Can we capture the intersections? Older black women, education, and health. *Women's Health Issues*, *22*(1), e91–e98. https://doi.org/10.1016/j.whi.2011.08.002

Hopton, J., Cree, A., Thompson, S., Jones, R., & Jones, R. (2018). An evaluation of the quality of HCR-20 risk formulations: A comparison between HCR-20 version 2 and HCR-20 version 3. *International Journal of Forensic Mental Health*, *17*(2), 195–201. https://doi.org/10.1080/14999013.2018.1460424

Howner, K., Andiné, P., Bertilsson, G., Hultcrantz, M., Lindström, E., Mowafi, F., Snellman, A., & Hofvander, B. (2018). Mapping systematic reviews on forensic psychiatric care: A systematic review identifying knowledge gaps. *Frontiers in Psychiatry*, *9*, 452. https://doi.org/10.3389/fpsyt.2018.00452

Huang, Y., Kotov, R., de Girolamo, G., Preti, A., Angermeyer, M., Benjet, C., Demyttenaere, K., de Graaf, R., Gureje, O., Karam, A. N., Lee, S., Lépine, J. P., Matschinger, H., Posada-Villa, J., Suliman, S., Vilagut, G., & Kessler, R. C. (2009). DSM–IV personality disorders in the WHO world mental health surveys. *British Journal of Psychiatry*, *195*(1), 46–53. https://doi.org/10.1192/bjp.bp.108.058552

Hughes, L. A. (2005). The representation of females in criminological research. *Women & Criminal Justice*, *16*(1–2), 1–28. https://doi.org/10.1300/J012v16n01_01

Institute for Criminal Policy Research. (2022). *World female prison population up by 60% since 2000*. https://shorturl.at/cuwHX

Iqbal, M., Bardwell, H., & Hammond, D. (2021). Estimating the global economic cost of violence: Methodology improvement and estimate updates. *Defence and Peace Economics*, *32*(4), 403–426. https://doi.org/10.1080/10242694.2019.1689485

Jeandarme, I., Pouls, C., Oei, T. I., & Bogaerts, S. (2017). Forensic psychiatric patients with comorbid psychopathy: Double trouble? *International Journal of Forensic Mental Health*, *16*(2), 149–160. https://doi.org/10.1080/14999013.2017.1286414

* Lieser, K., & Rossdale, S. V. (2023). Predictive validity of the HCR-20v3 compared to the HCR-20v3 with the FAM in women. *International Journal of Forensic Mental Health*, *22*(3), 187–198. https://doi.org/10.1080/14999013.2022.2130478.

Lin, H. (2020). *Convert between different effect sizes*. Effect Size Converter. https://www.escal.site/

Logan, C. (2014). The HCR-20 version 3: A case study in risk formulation. *International Journal of Forensic Mental Health*, *13*(2), 172–180. https://doi.org/10.1080/14999013.2014.906516

Mann, C. R. (1989). Minority and female: A criminal justice double bind. *Social Justice*, *16*(4), 95–114. https://www.jstor.org/stable/29766503

* Mastromanno, B., Brookstein, D. M., Ogloff, J. R., Campbell, R., Chu, C. M., & Daffern, M. (2018). Assessing change in dynamic risk factors in forensic psychiatric inpatients: Relationship with psychopathy and recidivism. *Journal of Forensic Psychiatry & Psychology*, *29*(2), 323–336. https://doi.org/10.1080/14789949.2017.1377277

Merone, L., Tsey, K., Russell, D., & Nagle, C. (2022). Sex inequalities in medical research: A systematic scoping review of the literature. *Women's Health Reports*, *3*(1), 49–59. https://doi.org/10.1089/whr.2021.0083

Monjazeb, S., & Douglas, K. S. (2022). Do common risk factors for violence generalize across ethnicity? *Psychology Crime & Law*, *30*(5), 1–30. https://doi.org/10.1080/1068316X.2022.2097231

Montford, K. S., & Hannah-Moffat, K. (2021). The veneers of empiricism: Gender, race and prison classification. *Aggression & Violent Behavior*, *59*, 101475. https://doi.org/10.1016/j.avb.2020.101475

Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., Currie, M., Qureshi, R., Mattis, P., Lisy, K., & Mu, P.-F. (2020). Chapter 7: Systematic reviews of etiology and risk. In E. Aromataris & Z. Munn (Eds.), *JBI manual for evidence synthesis*. JBI. https://shorturl.at/zLNRW

NHS. (2018). *Adult psychiatric morbidity survey: Survey of mental health and wellbeing, England, 2014*. Nationalarchives.gov.uk. https://shorturl.at/bqzE2

Ogonah, M. G., Seyedsalehi, A., Whiting, D., & Fazel, S. (2023). Violence risk assessment instruments in forensic psychiatric populations: A systematic review and meta-analysis. *The Lancet Psychiatry*, *10*(10), 780–789. https://doi.org/10.1016/S2215-0366(23)00256-0

O'Shea, L. E., Mitchell, A. E., Picchioni, M. M., & Dickens, G. L. (2013). Moderators of the predictive efficacy of the historical, clinical and risk management-20 for aggression in psychiatric facilities: Systematic review and meta-analysis. *Aggression & Violent Behavior*, *18*(2), 255–270. https://doi.org/10.1016/j.avb.2012.11.016

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S. . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105906. https://doi.org/10.1016/j.ijsu.2021.105906

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews. A Product from the ESRC Methods Programme version. 1*(1), b92. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ed8b23836338f6fdea0cc55e161b0fc5805f9e27

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*, 615–620. https://doi.org/10.1007/s10979-005-6832-7

Rossdale, S. V., Tully, R. J., & Egan, V. (2020). The HCR-20 for predicting violence in adult females: A meta-analysis. *Journal of Forensic Psychology Research and Practice*, *20*(1), 15–52. https://doi.org/10.1080/24732850.2019.1681875

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

Smith, K. J., O'Rourke, S., & Macpherson, G. (2020). The predictive validity of the HCR20V3 within Scottish forensic inpatient facilities: A closer look at key dynamic variables. *International Journal of Forensic Mental Health*, *19*(1), 1–17. https://doi.org/10.1080/14999013.2019.1618999

Smith, S. W., Morash, M., Walling, B. M., & Adams, E. A. (2020). Precursors to probation and parole agent communication patterns with female clients. *Communication Studies*, *71*(2), 203–225. https://doi.org/10.1080/10510974.2020.1725080

* Sorge, A., Borrelli, G., Saita, E., & Perrella, R. (2022). Violence risk assessment and risk management: Case-study of filicide in an Italian woman. *International Journal of Environmental Research & Public Health*, *19*(12), 6967. https://doi.org/10.3390/ijerph19126967

Steyerberg, E. W., Bleeker, S. E., Moll, H. A., Grobbee, D. E., & Moons, K. G. (2003). Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology*, *56*(5), 441–447. https://doi.org/10.1016/S0895-4356(03)00047-7

Streb, J., Lutz, M., Dudeck, M., Klein, V., Maaß, C., Fritz, M., & Franke, I. (2022). Are women really different? Comparison of men and women in a sample of forensic psychiatric inpatients. *Frontiers in Psychiatry*, *13*, 857468. https://doi.org/10.3389/fpsyt.2022.857468

Strengthening The Reporting of Observational Studies in Epidemiology. (2024). *STROBE checklist of items that should be included in reports of cohort studies*. https://www.strobe-statement.org/checklists/

Talari, K., & Goyal, M. (2020). Retrospective studies–utility and caveats. *The Journal of the Royal College of Physicians of Edinburgh*, *50*(4), 398–402. https://doi.org/10.4997/jrcpe.2020.409

Tarpey, E., Stephenson, Z., & Barker, R. (2023). Risk formulation in forensic practice: A review of the evidence. *Journal of Forensic Practice*, *25*(3), 243–262. https://doi.org/10.1108/JFP-12-2022-0064

Tully, S. M., Bucci, S., & Berry, K. (2023). "My life isn't my life, it's the systems": A qualitative exploration of women's experiences of day-to-day restrictive practices as inpatients. *Journal of Psychiatric & Mental Health Nursing*, *30*(1), 110–122. https://doi.org/10.1111/jpm.12855

Turanovic, J. J., Reisig, M. D., & Pratt, T. C. (2015). Risky lifestyles, low self-control, and violent victimization across gendered pathways to crime. *Journal of Quantitative Criminology*, *31*(2), 183–206. https://doi.org/10.1007/s10940-014-9230-9

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence, V2*. Mental Health, Law, and Policy Institute, Simon Fraser University.

Wells, G., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2014). *Newcastle-Ottawa quality assessment scale cohort studies*. University of Ottawa. https://www.ncbi.nlm.nih.gov/books/NBK99082/bin/appb-fm4.pdf

Whiting, D., Lichtenstein, P., & Fazel, S. (2021). Violence and mental disorders: A structured review of associations by individual diagnoses, risk factors, and risk assessment. *The Lancet Psychiatry*, *8*(2), 150–161. https://doi.org/10.1016/S2215-0366(20)30262-5

* Wolf, V., Mayer, J., Steiner, I., Franke, I., Klein, V., Streb, J., & Dudeck, M. (2023). Risk factors for violence among female forensic inpatients with schizophrenia. *Frontiers in Psychiatry*, *14*, 1203824. https://doi.org/10.3389/fpsyt.2023.1203824

Yasrebi De Kom, F. M., Dirkzwager, A. J., Van Der Laan, P. H., & Nieuwbeerta, P. (2022). The effect of sanction severity and its interaction with procedural justice. *Criminal Justice & Behavior*, *49*(2), 200–219. https://doi.org/10.1177/00938548211038358