

University of Groningen

Gender-Sensitive Violence Risk Assessment

de Vogel, Vivienne; Bruggeman, Mieke; Lancel, Marike

Published in:
Criminal Justice and Behavior

DOI:
[10.1177/0093854818824135](https://doi.org/10.1177/0093854818824135)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

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 and Behavior*, 46(4), 528-549. <https://doi.org/10.1177/0093854818824135>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

GENDER-SENSITIVE VIOLENCE RISK ASSESSMENT

Predictive Validity of Six Tools in Female Forensic Psychiatric Patients

VIVIENNE DE VOGEL 

*De Forensische Zorgspecialisten
HU University of Applied Sciences Utrecht*

MIEKE BRUGGEMAN

Netherlands Institute for the Study of Crime and Law Enforcement

MARIKE LANCEL

*GGZ Drenthe
University of Groningen*

Most violence risk assessment tools have been validated predominantly in males. In this multicenter study, the Historical, Clinical, Risk Management–20 (HCR-20), Historical, Clinical, Risk Management–20 Version 3 (HCR-20^{V3}), Female Additional Manual (FAM), Short-Term Assessment of Risk and Treatability (START), Structured Assessment of Protective Factors for violence risk (SAPROF), and Psychopathy Checklist–Revised (PCL-R) were coded on file information of 78 female forensic psychiatric patients discharged between 1993 and 2012 with a mean follow-up period of 11.8 years from one of four Dutch forensic psychiatric hospitals. Notable was the high rate of mortality (17.9%) and readmission to psychiatric settings (11.5%) after discharge. Official reconviction data could be retrieved from the Ministry of Justice and Security for 71 women. Twenty-four women (33.8%) were reconvicted after discharge, including 13 for violent offenses (18.3%). Overall, predictive validity was moderate for all types of recidivism, but low for violence. The START Vulnerability scores, HCR-20^{V3}, and FAM showed the highest predictive accuracy for all recidivism. With respect to violent recidivism, only the START Vulnerability scores and the Clinical scale of the HCR-20^{V3} demonstrated significant predictive accuracy.

Keywords: risk assessment; gender; female forensic psychiatric patients; recidivism; mortality

Females represent only a minority in the criminal justice system. However, there is widespread agreement that female offending is on the rise, especially violent offending and

AUTHORS' NOTE: *The authors wish to thank Gerjonne Akkerman-Bouwsema, Anouk Bohle, Yvonne Bouman, Nienke Epskamp, Susanne de Haas, Loes Hagenauw, Paul ter Horst, Stéphanie Klein Tuente, Jeantine Stam, Eva de Spa, Nienke Verstegen, Michiel de Vries Robbé, and all other colleagues from FPC Oldenkotte, FPC Woenselse Poort, FPK Assen, and the Van der Hoeven Kliniek who contributed to this study. This study was conducted with official permission from the directors of the four settings. Correspondence concerning this article should be addressed to Vivienne de Vogel, Department of Research, De Forensische Zorgspecialisten, P.O. Box 174, 3500 AD Utrecht, The Netherlands; e-mail: vdevogel@dfzs.nl.*

CRIMINAL JUSTICE AND BEHAVIOR, 2019, Vol. 46, No. 4, April 2019, 528–549.

DOI: 10.1177/0093854818824135

Article reuse guidelines: sagepub.com/journals-permissions

© 2019 International Association for Correctional and Forensic Psychology

particularly among young women (Miller, Malone, & Dodge, 2010; Moretti, Catchpole, & Odgers, 2005). Hence, a disproportionate increase in the number of females entering the penitentiary system and forensic mental health care has been observed in many countries (for reviews, see Nicholls, Cruise, Greig, & Hinz, 2015; Odgers, Moretti, & Reppucci, 2005; Walmsley, 2015). In addition, it should be noted that the official prevalence rates of female offending might constitute an underestimation, as women usually commit less reported and less visible offenses, such as intimate partner violence (Desmarais, Reeves, Nicholls, Telford, & Fiebert, 2012) and child abuse (May-Chahal & Cawson, 2005). Considering the rising numbers of females entering forensic mental health care, it is important to know whether commonly applied tools and treatment methods are applicable to females, as the majority were developed for male forensic populations.

The use of violence risk assessment tools has become common practice in forensic mental health care. Structured violence risk assessment provides insight into risk and protective factors and offers guidelines for risk management and treatment and is thus of great importance for both society and the individual patient/offender. Prevention of (violent) offending by women is of utmost importance, also for the next generations. Children of violent or antisocial mothers have elevated risks of problems relating to mental and physical health, substance use, school, and offending behavior (Kim, Capaldi, Pears, Kerr, & Owen, 2009). In the past three decades, major progress has been made in the area of structured violence risk assessment, and multiple instruments have become available to guide mental health professionals in this process. However, the majority of risk assessment instruments are based on violence risk research conducted primarily in male samples. Moreover, research into the psychometric properties of these tools has been carried out mostly with men.

There is an ongoing debate on whether risk assessment procedures and tools are gender-neutral or gender-responsive (Salisbury, Boppre, & Kelly, 2016; Yesberg, Scanlan, Hanby, Serin, & Polaschek, 2015). Some scholars have taken the position that there is no reason to assume that male-based instruments do not apply to women because most risk factors are considered valid for both genders (see Rettinger & Andrews, 2010). Previous violent behavior, young age at first violent offense, and substance abuse have all been found to be valid risk factors for both females and males (Andrews et al., 2012). Other scholars have advocated the acknowledgment of unique risk factors and pathways to offending for females, for instance, relating to traumatic experiences and mental health issues (Blanchette & Brown, 2006; Brennan, Breitenbach, Dieterich, Salisbury, & van Voorhis, 2012; Davidson & Chesney-Lind, 2009; Salisbury et al., 2016). Some risk factors that have been found specifically in female populations are pregnancy at a young age (Messer, Maughan, Quinton, & Taylor, 2004), prostitution (Morgan & Patton, 2002), and self-harm (Völlm & Dolan, 2009), although the predictive accuracy of these factors needs to be demonstrated more firmly. Mental health professionals who work with women on a daily basis recognize gender differences in violent offending and have called for more gender-sensitive assessment of violence risk factors (Adams & Freeman, 2002; Odgers et al., 2005). Despite the ongoing debate and the need for more research, it appears that the majority of risk factors are valid for both genders, but some risk factors seem to have a stronger or different impact on females compared with males. A distinction can be made between factors to which women are exposed more often (e.g., sexual victimization) and factors to which women are more sensitive, that is, factors that have a stronger effect on later violent or criminal behavior on women than men (e.g., disruptions in social relationships; de Vogel & Nicholls, 2016).

Furthermore, it has been suggested that girls and women respond differently to protective factors compared with boys and men. For example, close family ties, positive social relationships, sound finances, and being religious appeared to have a stronger protective effect on females than males (Hart, O'Toole, Price-Sharps, & Shaffer, 2007; Rodermond, Kruttschnitt, Slotboom, & Bijleveld, 2016). Another factor that is considered as having potentially strong protective influences in females is self-efficacy (van Voorhis, Wright, Salisbury, & Bauman, 2010).

Summarizing, although there are many similarities for female and male offenders, several important gender differences in risk and protective factors for (violent) offending have been found, and it can therefore be questioned whether commonly used risk assessment tools are sufficiently valid and useful in female populations.

THE VALUE OF RISK ASSESSMENT TOOLS IN FEMALES

Research has revealed that unstructured clinical judgment of violence risk is sensitive to sex-based biases. Mental health professionals of both genders tend to underestimate the risk of violence in female psychiatric patients (Skeem et al., 2005), and the use of structured tools is recommended. Most commonly used structured risk assessment tools can be divided into actuarial tools and tools according to the Structured Professional Judgment (SPJ) method. An important distinction between actuarial and SPJ tools is how the final judgment is established. In actuarial tools, the risk conclusion is obtained via an algorithm, whereas in SPJ tools information should be integrated, combined, and weighed by the assessor to come to a structured, individualized final judgment. An example of a widely used actuarial tool for general offending is the Level of Service Inventory (LSI, or revised versions; see Andrews & Bonta, 2000). A large body of research has demonstrated good predictive accuracy for the LSI in both male and female populations (Olver, Stockdale, & Wormith, 2014; though for a critical discussion about the use of the LSI in females, see Salisbury et al., 2016). One of the most widely used SPJ risk assessment instruments for violence in forensic psychiatry is the Historical, Clinical, Risk Management–20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997). Several studies found lower predictive validity for females than males (see the reviews of Garcia-Mansilla, Rosenfeld, & Nicholls, 2009; McKeown, 2010). Yet, a meta-analysis showed that the HCR-20 had the best predictive efficacy among samples containing higher proportions of women, patients with schizophrenia, and Caucasians (O'Shea, Mitchell, Picchioni, & Dickens, 2013). The Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009) is another commonly used SPJ tool, which is fully dynamic and assesses different forms of risk. Not many studies have yet been conducted on the value of the START in female populations, but the results so far have shown good predictive accuracy in women (O'Shea & Dickens, 2015; Viljoen, Nicholls, Greaves, de Ruiter, & Brink, 2011). Predictive accuracy of SPJ tools assessing protective factors was found to be lower in females compared with males (Viljoen et al., 2016). In the Structured Assessment of Protective Factors for violence risk (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbé, 2012), the most accurate predictors for abstention from violence differed. For men, the protective factors self-control, work, and attitudes toward authority were the strongest predictors for not committing violent incidents during treatment, whereas for women, leisure activities, coping, and intelligence were the strongest (de Vries Robbé, de Vogel, Wever, Douglas, & Nijman, 2016).

Psychopathy is considered to be a strong predictor of (violent) recidivism and is therefore often incorporated in risk assessment tools, such as the HCR-20 (Leistico, Salekin, DeCoster, & Rogers, 2008). The Psychopathy Checklist–Revised (PCL-R; Hare, 2003) is widely used for the assessment of psychopathy, and although it is not a risk assessment instrument, PCL-R total scores have proven to be predictive for violence (Leistico et al., 2008). Most studies into the value of the PCL-R have been conducted in male samples. Studies in female samples have yielded mixed results, with some finding similar results as for males, whereas others found lower predictive accuracy in females compared with males (for discussions, see Logan, Weizmann-Henelius, 2012; Nicholls, Ogloff, Brink, & Spidel, 2005). Overall, the PCL-R is assumed to have relevance in violence risk assessment in female offenders (Nicholls et al., 2005), yet concerns have been expressed about whether the PCL-R satisfactorily captures the construct of psychopathy in women (Forouzan & Cooke, 2005).

Garcia-Mansilla et al. (2009) reviewed the literature on different methods of violence risk assessment in a range of female populations. The conclusion of this review was that although structured methods of risk assessment are more accurate in predicting violent behavior than unstructured ones, the research supporting the applicability of violence risk assessment tools in female populations remains equivocal. McKeown (2010) also conducted a literature review on violence risk assessment and psychopathy in women and concluded that more research is needed with a particular focus on additional risk factors for women. More recently, a systematic review of nine risk assessment tools used in 15 studies including adult female offenders was carried out by Geraghty and Woodhams (2015). The findings from this review indicate that none of the measures demonstrated strong predictive validity in female populations. The LSI was found to be the most effective tool for assessing both violent and general recidivism in women.

GENDER-SENSITIVE RISK ASSESSMENT TOOLS

To date, only a few risk assessment instruments are available that take gender into account. Some actuarial risk assessment tools have been developed or adapted for use with female offenders. For example, the Women's Risk Needs Assessments (WRNA; van Voorhis et al., 2010) are actuarial tools that assess both gender-neutral (e.g., substance abuse, criminal history, financial problems) and gender-responsive (e.g., sexual trauma, mental health issues, relationship conflict) factors in female prisoners. Research found support for the predictive accuracy of the WRNA in female prison populations in the United States (van Voorhis, Bauman, & Brushett, 2013). To our knowledge, these tools have not yet been validated for forensic psychiatric patients. Furthermore, two SPJ tools specifically for females are available. The Early Assessment Risk List for Girls (EARL-21G; Levene et al., 2001) is an SPJ instrument for antisocial behavior in girls aged between 6 and 12 years. The Female Additional Manual (FAM; de Vogel, de Vries Robbé, van Kalmthout, & Place, 2014) is a gender-sensitive risk assessment guideline for adult female (forensic) psychiatric patients that has been developed as an addition to the HCR-20, or its revision, the Historical, Clinical, Risk management–20 Version 3 (HCR-20^{V3}; Douglas, Hart, Webster, & Belfrage, 2013). The FAM contains eight additional risk factors (e.g., prostitution, low self-esteem) and three additional final risk judgments (self-destructive behavior, victimization, and nonviolent criminal behavior). There is still relatively little empirical evidence for these SPJ gender-sensitive tools. It should be noted that research on female offenders, specifically forensic psychiatric patients, and predicting violent behavior is challenging because of small sample

sizes and relatively high rates of chronic psychiatric admission. Consequently, such studies require collaboration between settings and a longer research period (see also Burman, Batchelor, & Brown, 2001).

THE CURRENT STUDY

To improve current risk assessment and risk management practices, our knowledge and understanding of female offenders and their risk of recidivism need to be enlarged. The current study aims to evaluate the predictive accuracy of multiple tools for recidivism, most importantly violent recidivism, in a female forensic population. The predictive validity of five SPJ risk assessment instruments—the HCR-20, HCR-20^{V3}, FAM, START, and SAPROF—as well as the PCL-R for violent and general recidivism was examined in a sample of 78 women discharged from four Dutch forensic psychiatric settings. In addition, we examined discriminant validity by comparing risk scores between recidivists and nonrecidivists and convergent validity by analyzing Pearson correlations coefficients between the different instruments and their subscales. This study was mainly explorative, as the literature is still inconclusive on risk assessment in female forensic psychiatric patients. As the FAM has been developed as an addition to the HCR-20/HCR-20^{V3} specifically for female offenders, it is hypothesized that the FAM will show higher predictive accuracy (area under the curve [AUC] values) than these two tools for violence.

METHOD

PARTICIPANTS

The majority of the 78 women included in the present study were of Dutch descent ($n = 68$, 87.2%). The mean age at the time of admission to the forensic psychiatric hospitals was 33.7 years ($SD = 9.6$, range = 20-65) and of discharge, 38.6 years ($SD = 9.7$, range = 22-69). Most of the women had been involuntarily admitted by a court order ($n = 71$, 91.0%), called TBS-order (terbeschikkingstelling; translated as “detained under a treatment order”). The TBS-order is imposed by court on offenders who have committed a serious violent offense, are considered to be at high risk of re-offending, and who have diminished responsibility for the offense because of severe psychopathology. Offenses for which the women were admitted to forensic psychiatry were homicide offenses ($n = 42$, 53.8%), arson ($n = 18$, 23.1%), violent offenses ($n = 11$, 14.1%), property offenses ($n = 5$, 6.4%), and sexual offenses ($n = 2$, 2.6%). The mean treatment duration of the women was 62.4 months ($SD = 41.9$, range = 1-208). The majority had been physically, sexually, and/or emotionally abused in their childhood ($n = 57$, 73.1%) and/or adult life ($n = 50$, 64.1%). Furthermore, substance abuse ($n = 51$, 65.4%), borderline personality disorder (BPD; $n = 40$, 51.1%), and traits of BPD ($n = 19$, 24.4%) were highly prevalent in this sample (for more, see de Vogel, Stam, Bouman, Ter Horst, & Lancel, 2016).

MEASURES

HCR-20/HCR-20^{V3}

The HCR-20 contains 10 risk factors from the past (Historical scale), five from the present (Clinical scale), and five relating to the future (Risk Management scale). After coding the presence of the 20 items on a 3-point scale (0, 1, 2), the risk factors should not simply

be summed up, but interpreted, integrated, and weighed by the assessor to make a final risk judgment (Low, Moderate, High; in the present study coded on a 5-point level: 1 = *low*; 2 = *low to moderate*; 3 = *moderate*; 4 = *moderate to high*; 5 = *high*). Most studies with the HCR-20 found acceptable interrater reliability and validity (Douglas et al., 2017). However, the majority of these studies were conducted in male populations, and research in female populations revealed mixed results (McKeown, 2010). The HCR-20^{V3} is the revised version of the HCR-20 and still contains 20 items divided into the Historical, Clinical, and Risk Management scales, but the content of items has changed and several items include sub-items. The items should be rated on a three-level response format (No, Partially, Yes; in the present study transposed to numerical scores 0, 1, 2). To date, little is known about the predictive validity of the HCR-20^{V3} for women. A small-scale study with insanity acquittals showed that the relationship between scale scores and violence was stronger among men than women, although gender was not a significant moderator in logistic regression analyses predicting likelihood of violence (Green et al., 2016).

FAM

The FAM was developed as an additional guideline to the HCR-20 for assessing risk of violence among women. In 2014, it was adapted for use with the HCR-20^{V3}. The FAM contains additional guidelines to two historical items of the HCR-20^{V3} (personality disorder and traumatic experiences) and eight new items with specific relevance to women (prostitution, parenting difficulties, pregnancy at young age, suicide attempt/self-harm, covert/manipulative behavior, low self-esteem, problematic child care responsibility, and problematic future intimate relationship). Furthermore, three additional final risk judgments can be coded with the FAM (the risk for self-destructive behavior, victimization, and nonviolent criminal behavior). With respect to the psychometric properties of the FAM, only preliminary results from small samples are available. They show good interrater reliability and moderate to good predictive validity for inpatient violence, threatening behavior, and self-harm during treatment (Campbell & Beech, 2018; de Vogel et al., 2014; Greig, 2014; Griswold et al., 2016).

START

The START is an SPJ tool for short-term assessment of violence risk and treatability. It consists of 20 dynamic factors that can be coded both as vulnerability and as strength on a 3-point scale (0, 1, 2). In the START, multiple risks can be assessed next to the risk of violence to others: risk of suicide, self-harm, self-neglect, unauthorized absence, substance use, and risk of being victimized. O'Shea and Dickens (2015) examined the predictive validity of the START in a sample of secure psychiatric patients using START codings by multidisciplinary teams of mental health professionals. They found the START to be a stronger predictor of aggression and self-harm in women than men. In a Canadian study, significant predictive accuracy was found for both the Vulnerability and Strength scores of the START in a group of 48 female forensic patients (Viljoen et al., 2011).

SAPROF/Protective Factors

The SAPROF is an instrument to assess protective factors for violence risk in adults that should always be used in addition to risk-focused risk assessment tools, such as the

HCR-20/HCR-20^{V3}. The tool includes three scales: Internal (five items), Motivational (seven items), and External (five items). Research so far has demonstrated good psychometric properties for the SAPROF (see de Vries Robbé et al., 2016), although mainly examined in male patients. Results for female patients were less strong. Viljoen et al. (2016) examined protective factors measured with different tools (SAPROF, START) in civil psychiatric patients and concluded that SPJ tools utilizing both risk and protective factors performed better. Gender was found to be a moderator in predicting severe aggression when using risk assessment tools. Both the SAPROF and the START Strength scores demonstrated superior predictive validity for men compared with women.

PCL-R

The PCL-R was developed to assess psychopathy and comprises 20 factors coded on a 3-point scale (0, 1, 2) that are divided into four facets: Interpersonal (e.g., conning, manipulative, glibness), Affective (e.g., shallow affect, callous), Impulsive (e.g., lack of goals, irresponsibility), and Antisocial (e.g., criminal versatility, poor behavioral control). Several studies have been conducted on predictive accuracy in female samples (e.g., Richards, Casey, & Lucente, 2003), and the results have been mixed (for discussions, see Logan & Weizmann-Henelius, 2012; Nicholls et al., 2005). In a previous study within the current research project, the predictive validity of the PCL-R for physical violence during treatment was found to be good for men and moderate for women. When verbal violence was included in the definition of violence during treatment, the predictive validity of the PCL-R was good for both genders (de Vogel & Lancel, 2016).

PROCEDURE

In 2012, a multicenter research project was started on gender differences in forensic psychiatry (for more, see de Vogel et al., 2016). The historical items of the HCR-20, HCR-20^{V3}, and FAM were coded on file information for all women who have been admitted to one of four different forensic psychiatric facilities in the Netherlands between 1984 and 2013 ($N = 297$). This sample represents nearly all female forensic psychiatric patients in the Netherlands in that period. In addition, an extensive list of criminal, demographic, psychiatric, and treatment characteristics was coded. Overall, file information was extensive and contained abundant collateral information (e.g., police reports, psychological reports, treatment plans, and evaluations). Each researcher rated the quality of the file information on a 0 (*insufficient*) to 100 (*excellent*) scale, based on availability of reliable information about the entire lifespan and from multiple sources. When two or more items could not be coded for the historical items of the risk assessment instruments because of lack of information, the quality of the file was coded as insufficient (i.e., below 50). All files with a score below 50 ($n = 17$) were excluded from the analysis, resulting in a sample of 280 women.

All female patients who had been discharged with a follow-up period of at least 3 years in society were included in the present study ($N = 78$, 27.9%, discharged between 1993 and 2012).¹ The quality of these 78 files was generally judged as good, with a mean score of 75.6 ($SD = 12.3$, range = 50-100). The remaining cases ($n = 202$, 72.1%) were still in forensic or general psychiatric care or were only recently discharged. We chose the period of 3 years because it usually takes quite some time before an offender is officially convicted and before these convictions can be retrieved from the official judicial documentation

register. For these 78 cases, we also coded the dynamic items of the HCR-20, HCR-20^{V3}, and FAM, including the final risk judgments, as well as the SAPROF and START based on the most recent information available in the files. The six tools were chosen because these are the standardly used and validated instruments in the participating settings. Since 2005, the HCR-20, including the PCL-R score, is one of two instruments mandated in the Netherlands to use with TBS patients. In 2013, the HCR-20 was replaced by the HCR-20^{V3}. The majority of the PCL-R ratings were retrieved from the hospital files, meaning that they were usually coded based on interviews plus file information by diagnosticians in the participating hospitals (see de Vogel & Lancel, 2016). The other tools were coded by a group of 10 trained and experienced researchers (psychologists and criminologists) who were all blind to outcome data. When four or more items could not be coded, the quality was considered insufficient. Two cases had more than four missing items on the START and were left out of the predictive validity analyses for the START. For the other cases, the maximum number of missing items in total of the SPJ tools was three. The missing items were prorated by applying the mean score on the item of the total group.

Interrater reliability of the HCR-20/FAM Historical items was previously established for 25 of the 275 women in this research project and found to be good (intraclass correlation [ICC] = .93, $p < .001$, see de Vogel et al., 2016). For the other tools, it was not possible to test interrater reliability due to practical reasons, but previous studies in Dutch forensic psychiatric samples have found good interrater reliability for the PCL-R (Hildebrand, de Ruiter, de Vogel, & van der Wolf, 2002), SAPROF (de Vries Robbé et al., 2016), HCR-20 (de Vogel, 2005), and HCR-20^{V3} (de Vogel, van den Broek, & de Vries Robbé, 2014).

Official reconvection data were retrieved in December 2016 from the Judicial Documentation register of the Ministry of Justice and Security. For seven women, recidivism data could not be obtained, and these cases were excluded from the predictive validity analyses. Of these seven women, four were deceased and three could not be retrieved in the system, possibly due to the fact that they initially had conditional convictions because of minor index offenses. In total, 14 of the 78 women (17.9%) were deceased after discharge. As 10 of them had had a follow-up period of at least 3 years in society, these cases were not excluded from predictive analyses. The total sample for the analyses of predictive validity for any and violent recidivism consisted of 71 women. The mean follow-up period of these women was 11.8 years ($SD = 4.9$, range = 5.8-23.0). Predictive validity was analyzed for both a preset follow-up period of 3 years and for the total mean follow-up period of 11.8 years.

DATA ANALYSIS

Statistical analyses were carried out using SPSS Version 25. Receiver operating characteristic (ROC) analyses were conducted to assess predictive validity of the six tools for violent or general recidivism after discharge. ROC analyses result in AUC values. ROC analyses are reasonably unaffected by base rates, and although there are limitations, most research into risk assessment apply ROC analyses, facilitating comparison (Flores, Holsinger, Lowenkamp, & Cohen, 2017; Mossman, 2013). Rice and Harris (2005) provided guidelines for interpreting AUC values and facilitated comparison across studies by applying different effect sizes. AUC values between .56 and .64 can be compared with Cohen's d of .20 and are interpreted as low effect; AUC values between .64 and .71 with Cohen's d of .50, medium effect; and AUC values of .71 and above with Cohen's d of .80, high effect.

RESULTS

OUTCOME

Nine of the 78 women (11.5%) were readmitted to civil psychiatry after having lived in society for a period of time. Fourteen women (17.9%) were deceased after discharge at a mean age of 44.6 years ($SD = 9.2$, range = 29-59). We were not able to retrieve the causes of death for all of these women, but at least three women died by suicide. Twenty-four of the 71 women whose recidivism data could be retrieved (33.8%) were officially reconvicted for one or more offenses. The mean number of reconvictions for these 24 women was 5.5 ($SD = 4.7$, range = 1-18). Thirteen (18.3%) women were reconvicted for violent offense(s), including attempted homicide ($n = 2$), arson ($n = 5$), threat ($n = 6$), assault ($n = 6$), and stalking ($n = 1$). Eleven women (15.5%) were convicted because of nonviolent offenses, such as property offenses and fraud. Nine women (12.7%) were convicted for both violent and nonviolent offenses.

DISCRIMINANT VALIDITY

To examine discriminant validity, we compared the mean scores on the instruments for the recidivists and nonrecidivists for all recidivism and for violent recidivism. With respect to all recidivism, Table 1 shows that the mean scores on the FAM, HCR-20^{V3} (and their combination), and START differed significantly between recidivists and nonrecidivists. Table 1 also displays that with respect to violent recidivism, only the mean START Vulnerability scores differed significantly between women with a violent recidivism and women who had not recidivated with a violent offense.

CONVERGENT VALIDITY

Table 2 presents the Pearson correlation coefficients between the different instruments, including the subscales and facets. All instruments' total scores significantly correlated with each other, suggesting that they measure a common construct of risk and have shared variance. In addition, almost all subscales and facet scores correlated significantly with each other and the total scores. An exception was the correlation between PCL-R Facet 2 and HCR-20^{V3} Risk Management scale and most of the correlations with PCL-R Facet 1 Interpersonal and the SAPROF External scale.

PREDICTIVE VALIDITY RECIDIVISM

Table 3 shows that the total scores of the FAM, HCR-20^{V3}, PCL-R, and START Vulnerability demonstrated significant predictive accuracy for all recidivism (including violence) for the preset follow-up period of 3 years. In addition, the combination FAM and HCR-20^{V3}, the HCR-20 Clinical scale, the HCR-20^{V3} Historical and Clinical scales, the PCL-R Facet 1 Interpersonal, and the FAM final risk judgment nonviolent criminal behavior were significant predictors for all recidivism. With respect to the longer follow-up period of 11.8 years, the FAM, the HCR-20^{V3}, the combination FAM and HCR-20^{V3}, the HCR-20 and HCR-20^{V3} Clinical scales, the FAM final risk judgment nonviolent criminal behavior, and START Vulnerability remained significant predictors. Predictive validity was also analyzed specifically for nonviolent recidivism ($n = 11$) for the FAM final risk judgment nonviolent criminal behavior: $AUC = .86$, $SE = .061$, 95% confidence interval

TABLE 1: Mean Total Scores (SD) for Those With and Without Recidivism

Recidivism: Any offense	Yes (<i>n</i> = 24)	No (<i>n</i> = 47)	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
FAM	16.2 (4.6)	13.8 (4.2)	-2.2*	.031	.544
FAM + HCR-20 ^{V3}	31.1 (8.3)	26.0 (8.1)	-2.5*	.015	.622
HCR-20	24.4 (6.3)	22.2 (6.2)	-1.4	.161	.352
HCR-20 ^{V3}	22.5 (6.3)	18.8 (6.2)	-2.4*	.021	.591
PCL-R	16.9 (7.3)	13.7 (6.9)	-1.8	.073	.451
SAPROF	10.9 (5.4)	12.1 (5.3)	0.94	.351	.224
START Vulnerability	22.3 (7.8)	17.3 (7.8)	-2.5*	.016	.641
START Strength	16.2 (9.2)	19.2 (7.2)	1.4	.162	.363
Recidivism: Violent offense	Yes (<i>n</i> = 13)	No (<i>n</i> = 58)	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
FAM	15.7 (4.9)	14.4 (4.4)	-0.95	.346	.279
FAM + HCR-20 ^{V3}	31.5 (7.9)	26.9 (8.4)	-1.8	.075	.564
HCR-20	24.7 (6.5)	22.6 (6.2)	-1.1	.274	.330
HCR-20 ^{V3}	23.2 (5.8)	19.4 (6.4)	-1.9	.053	.622
PCL-R	15.3 (6.7)	14.7 (7.3)	-0.24	.939	.086
SAPROF	11.1 (5.6)	11.8 (5.4)	0.43	.670	.127
START Vulnerability	23.6 (6.1)	17.7 (8.1)	-2.5*	.017	.822
START Strength	15.5 (6.9)	18.9 (8.1)	1.4	.187	.452

Note. The follow-up period was 11.8 years. FAM = Female Additional Manual; HCR = Historical Clinical Risk; PCL-R = Psychopathy Checklist-Revised; SAPROF = Structured Assessment of Protective Factors; START = Short-Term Assessment of Risk and Treatability.

[CI] = [.743, .981], $p = .001$ for the 3-year follow-up period and $AUC = .64$, $SE = .109$, 95% CI = [.420, .849], $p = .157$ for the 11.8-year follow-up period.

Table 4 shows that none of the instruments were significant predictors for violent recidivism for the preset follow-up period of 3 years. The START Vulnerability and the HCR-20^{V3} Clinical scale demonstrated significant predictive validity for violent recidivism for the follow-up period of 11.8 years.

PREDICTIVE VALIDITY MORTALITY

Considering the unexpectedly high mortality rate at a relatively young mean age, we decided to conduct post hoc analyses to explore predictive validity of the six tools for mortality for the total follow-up period (mean = 10.9 years, $N = 78$). It was not possible to retrieve data on the causes of death, but early death may be related to self-destructive or suicidal behavior and psychopathology, especially BPD and substance use disorders (see Chesney, Goodwin, & Fazel, 2014). Although the instruments were not developed to assess risk of self-destructive or suicidal behavior, there is overlap in risk factors for violence to others and to the self (see de Vogel et al., 2014; Völlm & Dolan, 2009; Webster et al., 2009). The post hoc analyses showed that none of the instruments and subscales were significant predictors of mortality, with one exception, that is, PCL-R Facet 1 Interpersonal (see Table 5). This facet was a significant negative predictor of mortality, or in other words, a protective factor.

DISCUSSION

The present multicenter study is one of the first studies comparing the predictive validity of different SPJ risk assessment tools and the PCL-R for recidivism in a sample

TABLE 2: Pearson Correlation Coefficients for Total Scores and Subscale Scores of the Six Instruments

Instrument	FAM total	HCR-20 total	H-scale	C-scale	R-scale	HCR-20 ^{v3} total	H-scale	C-scale	R-scale	PCL-R total	F1	F2	F3	F4	SAPROF total	SAPROF Internal	SAPROF Motivational	SAPROF External	START V	START S
FAM total	—																			
HCR-20 total	.726**	—																		
H-scale	.715**	.863**	—																	
C-scale	.483**	.829**	.494**	—																
R-scale	.592**	.856**	.539**	.770**	—															
HCR-20 ^{v3} total	.664**	.943**	.743**	.847**	.860**	—														
H-scale	.617**	.839**	.834**	.629**	.605**	.879**	—													
C-scale	.553**	.820**	.499**	.898**	.832**	.875**	.596**	—												
R-scale	.516**	.762**	.439**	.756**	.886**	.827**	.500**	.800**	—											
PCL-R total	.587**	.585**	.590**	.487**	.362**	.598**	.659**	.447**	.345**	—										
F1	.298**	.168	.152	.196	.077	.208	.223	.177	.129	.662**	—									
F2	.289**	.393**	.298**	.490**	.247**	.398**	.425**	.394**	.166	.718**	.501**	—								
F3	.385**	.615**	.616**	.642**	.428**	.600**	.646**	.428**	.383**	.813**	.302**	.381**	—							
F4	.385**	.500**	.552**	.367**	.281**	.541**	.622**	.357**	.308**	.705**	.311**	.289*	.567**	—						
SAPROF total	-.477**	-.713**	-.542**	-.684**	-.643**	-.686**	-.640**	-.607**	-.491**	-.474**	-.196	-.467**	-.407**	-.313**	—					
Internal	-.321**	-.577**	-.415**	-.550**	-.562**	-.518**	-.469**	-.512**	-.343**	-.307**	-.053	-.334**	-.269**	-.299**	.755**	—				
Motivational	-.436**	-.679**	-.492**	-.689**	-.617**	-.683**	-.629**	-.629**	-.481**	-.474**	-.227	-.495**	-.400**	-.261*	.917**	.573**	—			
External	-.321**	-.299*	-.306**	-.211	-.212	-.271*	-.290*	-.124	-.243*	-.252*	-.137	-.151	-.222	-.158	.550**	.165	.313**	—		
START V	.605**	.777**	.480**	.783**	.809**	.795**	.552**	.818**	.775**	.418**	.162	.320**	.422**	.321**	-.486**	-.395**	-.540**	-.035	—	
START S	-.545**	-.771**	-.471**	-.788**	-.793**	-.790**	-.554**	-.805**	-.768**	-.447**	-.138	-.410**	-.435**	-.304*	.511**	.374**	.583**	.061	-.889**	—

Note. FAM = Female Additional Manual; HCR = Historical Clinical Risk; H-scale = Historical scale; C-scale = Clinical scale; R-scale = Risk Management scale; PCL-R = Psychopathy Checklist-Revised; F1 = Interpersonal; F2 = Affective; F3 = Impulsive; F4 = Antisocial; SAPROF = Structured Assessment of Protective Factors; START = Short-Term Assessment of Risk and Treatability; V = Vulnerability; S = Strength.
p* < .05. *p* < .01.

TABLE 3: Predictive Validity for All Recidivism After Discharge (N = 71)

Instrument	All recidivism 3 years, <i>n</i> = 14			All recidivism <i>M</i> = 11.8 years, <i>n</i> = 24		
	AUC	<i>SE</i>	95% CI	AUC	<i>SE</i>	95% CI
FAM total score	.676*	.082	[.516, .836]	.648*	.070	[.511, .786]
FAM + HCR-20	.630	.088	[.458, .801]	.598	.073	[.455, .741]
FAM + HCR-20 ^{V3}	.695*	.068	[.562, .829]	.661*	.069	[.525, .797]
HCR-20 total score	.632	.072	[.492, .771]	.590	.072	[.449, .732]
Historical scale	.537	.084	[.373, .701]	.515	.074	[.370, .660]
Clinical scale	.684*	.069	[.548, .819]	.662*	.070	[.525, .800]
Risk Management scale	.659	.065	[.531, .787]	.590	.072	[.448, .731]
HCR-20 ^{V3} total score	.711*	.059	[.595, .827]	.667*	.068	[.534, .800]
Historical scale	.672*	.066	[.542, .802]	.636	.070	[.498, .773]
Clinical scale	.680*	.066	[.550, .811]	.673*	.066	[.544, .802]
Risk Management scale	.667	.070	[.529, .804]	.641	.068	[.508, .774]
FRJ Violence	.537	.087	[.366, .707]	.597	.075	[.449, .744]
FRJ Non-violent (FAM)	.780**	.062	[.659, .901]	.679*	.067	[.547, .811]
PCL-R total score	.710*	.087	[.523, .863]	.601	.074	[.456, .746]
F1 Interpersonal	.724*	.078	[.571, .877]	.602	.071	[.462, .742]
F2 Affective	.642	.080	[.485, .784]	.566	.070	[.454, .737]
F3 Impulsive	.634	.076	[.485, .784]	.594	.072	[.454, .735]
F4 Antisocial	.620	.085	[.454, .786]	.595	.072	[.454, .737]
SAPROF total score	.391	.081	[.232, .550]	.434	.075	[.286, .581]
Internal scale	.403	.076	[.253, .553]	.473	.072	[.332, .613]
Motivational scale	.363	.073	[.221, .506]	.403	.072	[.261, .545]
External scale	.521	.080	[.364, .677]	.518	.074	[.373, .662]
START Vulnerability ^a	.748**	.061	[.626, .867]	.698*	.070	[.547, .822]
START Strength ^a	.364	.074	[.219, .508]	.394	.080	[.238, .551]

Note. The AUC values of the SAPROF and START Strength are reversed as these contain protective factors that aim to predict nonoffending. AUC = area under the curve; CI = confidence interval; FAM = Female Additional Manual; HCR = Historical Clinical Risk; FRJ = Final Risk Judgment; PCL-R = Psychopathy Checklist-Revised; SAPROF = Structured Assessment of Protective factors; START = Short-Term Assessment of Risk and Treatability.

^aTwo cases were left out of the analyses for the START Vulnerability and Strength because they had too many missing items (*n* = 69).

p* < .05. *p* < .01.

of discharged female forensic psychiatric patients. Overall, it can be concluded that the predictive validity for all types of recidivism was moderate, but low for violence. This is in accordance with previous studies (Coid et al., 2009), the conclusion of the systematic review of Geraghty and Woodhams (2015), and the reviews of McKeown (2010) and Garcia-Mansilla et al. (2009). In our female population, dynamic risk factors were the strongest predictors. The START Vulnerability, HCR-20^{V3}, especially the Clinical scale, and FAM total score and final risk judgment nonviolent criminal behavior demonstrated the strongest predictive accuracy for all recidivism for both medium- and long-term follow-up. For violent recidivism, only the START Vulnerability scores and the HCR-20^{V3} Clinical scale were significant predictors. The most important results and implications of this study will be discussed below.

TABLE 4: Predictive Validity for Violent Recidivism After Discharge (N = 71)

Instrument	Violent recidivism 3 years, <i>n</i> = 6			Violent recidivism <i>M</i> = 11.8 years, <i>n</i> = 13		
	AUC	SE	95% CI	AUC	SE	95% CI
FAM total score	.587	.128	[.336, .838]	.587	.094	[.402, .772]
FAM + HCR-20	.523	.123	[.282, .764]	.562	.093	[.380, .744]
FAM + HCR-20 ^{V3}	.619	.080	[.463, .775]	.651	.080	[.494, .808]
HCR-20 total score	.563	.090	[.387, .733]	.592	.087	[.421, .763]
Historical scale	.467	.118	[.235, .698]	.521	.096	[.330, .709]
Clinical scale	.549	.099	[.354, .743]	.641	.085	[.473, .808]
Risk Management scale	.628	.067	[.498, .759]	.595	.080	[.437, .753]
HCR-20 ^{V3} total score	.635	.071	[.496, .774]	.672	.076	[.523, .821]
Historical scale	.604	.085	[.437, .771]	.649	.083	[.486, .811]
Clinical scale	.655	.086	[.487, .775]	.690*	.074	[.544, .835]
Risk Management scale	.544	.084	[.380, .707]	.618	.075	[.472, .764]
FRJ Violence	.488	.138	[.217, .765]	.666	.091	[.487, .845]
PCL-R total score	.457	.131	[.201, .713]	.591	.084	[.315, .866]
F1 Interpersonal	.596	.110	[.381, .812]	.596	.080	[.179, .864]
F2 Affective	.503	.124	[.260, .745]	.495	.080	[.152, .693]
F3 Impulsive	.496	.095	[.310, .682]	.746	.087	[.560, .931]
F4 Antisocial	.473	.118	[.241, .704]	.655	.097	[.351, .959]
SAPROF total score	.503	.110	[.281, .714]	.476	.092	[.343, .704]
Internal scale	.477	.105	[.272, .682]	.537	.086	[.294, .631]
Motivational scale	.392	.088	[.220, .565]	.382	.079	[.463, .774]
External scale	.672	.098	[.480, .864]	.588	.097	[.221, .603]
START Vulnerability ^a	.697	.084	[.533, .860]	.704*	.071	[.565, .843]
START Strength ^a	.481	.073	[.338, .625]	.407	.080	[.250, .564]

Note. The AUC values of the SAPROF and START Strength are reversed as these contain protective factors that aim to predict nonoffending. AUC = area under the curve; CI = confidence interval; FAM = Female Additional Manual; HCR = Historical Clinical Risk; FRJ = Final Risk Judgment; PCL-R = Psychopathy Checklist-Revised; SAPROF = Structured Assessment of Protective Factors; START = Short-Term Assessment of Risk and Treatability.

^aTwo cases were left out of the analyses for the START Vulnerability and Strength because they had too many missing items (*n* = 69).

**p* < .05.

First, notable in this sample was the high rate of mortality and readmission to psychiatry after having lived in society for a period of time. Of the 280 women included in the multi-center research project, only 78 were discharged into society with a follow-up period of at least 3 years. Of those 78, 14 had died and nine were readmitted to civil psychiatry, leaving a total of 55 discharged women currently living in society. Moreover, recidivism data could not be retrieved for seven of these women. These findings indicate that female forensic psychiatric patients form a difficult sample to follow-up, which is in line with previous studies conducted in female forensic samples (Davies, Clarke, Hollin, & Duggan, 2007; Sahota et al., 2010). Still, in our study, four of the 14 (28.6%) women who had deceased within the follow-up period had committed a new offense before their death. These 14 women had died at a relatively young mean age of 44.6 years. We were not able to retrieve the causes of death for all of these women, but at least three women died by suicide. This

TABLE 5: Predictive Validity for Mortality After Discharge ($N = 78$)

Instrument	$M = 10.9$ years, $n = 14$		
	AUC	SE	95% CI
FAM total score	.550	.072	[.410, .691]
FAM + HCR-20	.571	.071	[.431, .710]
FAM + HCR-20 ^{V3}	.608	.068	[.475, .740]
HCR-20 total score	.601	.076	[.451, .751]
Historical scale	.598	.081	[.438, .757]
Clinical scale	.591	.081	[.432, .749]
Risk Management scale	.561	.078	[.409, .714]
HCR-20 ^{V3} total score	.607	.075	[.460, .754]
Historical scale	.605	.085	[.438, .772]
Clinical scale	.585	.079	[.430, .740]
Risk management scale	.600	.085	[.433, .767]
FRJ Self-destructive behavior (FAM)	.610	.084	[.446, .774]
PCL-R total score	.459	.079	[.304, .614]
F1 Interpersonal	.308*	.066	[.178, .437]
F2 Affective	.388	.072	[.246, .530]
F3 Impulsive	.567	.076	[.418, .715]
F4 Antisocial	.554	.092	[.374, .733]
SAPROF total score	.541	.078	[.475, .740]
Internal scale	.535	.082	[.375, .696]
Motivational scale	.503	.070	[.365, .641]
External scale	.502	.089	[.328, .675]
START Vulnerability ^a	.577	.077	[.424, .729]
START Strength ^a	.400	.078	[.250, .550]

Note. AUC = area under the curve; CI = confidence interval; FAM = Female Additional Manual; HCR = Historical Clinical Risk; FRJ = Final Risk Judgment; PCL-R = Psychopathy Checklist-Revised; SAPROF = Structured Assessment of Protective Factors; START = Short-Term Assessment of Risk and Treatability.

^aTwo cases were left out of the analyses for the START Vulnerability and Strength because they had too many missing items ($n = 76$).

* $p < .05$.

is not surprising considering the fact that BPD and substance use disorders, which were highly prevalent in the present sample, have been found as disorders with the highest suicide risk (Chesney et al., 2014). Possible explanations for early deaths besides suicide are the long-term physical consequences of severe self-harm and substance abuse (see, for example, Binswanger et al., 2007) and adverse childhood experiences that have been demonstrated to increase the risk of premature mortality (e.g., Brown et al., 2009). In the post hoc analyses, we found that PCL-R Facet 1 Interpersonal was a significant protective factor for mortality. An explanation could be that interpersonal psychopathic features such as conning and manipulative behavior, glibness, and grandiose sense of self-worth make these women less vulnerable for early death by severe self-destructive behavior including suicide. It should be emphasized that the results should be interpreted very carefully, as we did not have data on the causes of death. Not much research has been published on the relation between psychopathy and mortality, but in a recent study in a Finnish male sample a significant positive relationship was found between high PCL-R scores and premature death (Vaurio, Repo-Tiihonen, Kautiainen, & Tiihonen, 2018). The causes of death of the group with high PCL-R scores were mainly of violent nature or relating to a dangerous,

antisocial lifestyle, for example, homicides or accidents. The authors did not specify the results on the four facets. Our results are in contrast with these findings in men. An explanation may be the different manifestation of psychopathy in women—more subtle with less physical violence and a less antisocial, dangerous lifestyle compared with men. Our results are in line, though, with findings from a previous Dutch study, in which we found lower rates of self-destructive behavior during treatment for women scoring high on psychopathy versus women scoring low on psychopathy (de Vogel & Lancel, 2016). This study also found a lower incidence of physical violence during treatment for women scoring high on psychopathy versus women scoring low and men scoring high on psychopathy, which may confirm the suggestion above that psychopathy in females is manifested in a more subtle and less physically violent way.

Second, the recidivism rates of 36% for all offenses and 18% for violent offenses that were found in the present study are comparable to international findings in female prisoners (Coid et al., 2009; Rettinger & Andrews, 2010). It is more difficult to compare the recidivism rates with other forensic psychiatric samples, as not much has been published about recidivism of female forensic psychiatric samples. The scores of the recidivists and nonrecidivists differed significantly only for START Vulnerability, HCR-20^{V3}, and FAM.

Third, predictive validity was not very strong for most tools, with moderate accuracy for all recidivism, but low accuracy for violence. The observed AUC values for violent recidivism are lower than what is internationally found in studies with male samples (see, for example, Otto & Douglas, 2010). In previous studies in the Dutch forensic psychiatric population, predictive accuracy for male patients was considerably better for the PCL-R and HCR-20 (de Vogel & de Ruiter, 2005), the SAPROF (de Vries Robbé et al., 2016), and the HCR-20^{V3} (de Vogel et al., 2014) compared with the AUC values found for females in the current study. The START total risk score was a good predictor for all recidivism, especially for the 3-year follow-up period, but was also significant for long-term general and violent recidivism. This is notable because the START is fully dynamic and developed to assess short-term risks. Yet, these results are in line with previous studies that found good predictive accuracy of the START in female populations (O'Shea & Dickens, 2015; Viljoen et al., 2011). Strikingly, none of the other instruments showed significant predictive accuracy for violent recidivism. Only the Clinical scale of the HCR-20^{V3}, also consisting of only dynamic factors, was a significant predictor of violence. This significant predictive accuracy for the dynamic START and Clinical scale for the follow-up period of almost 12 years is remarkable, as typically predictive validities decrease over long follow-up periods. A possible explanation for the relatively strong predictive accuracy of dynamic factors in these female forensic psychiatric patients is their psychiatric background, particularly the high prevalence of BPD, often characterized by emotional instability and impulsivity. With respect to the PCL-R, the findings are in accordance with previous studies (see Logan, Weizmann-Henelius, 2012; McKeown, 2010). Good predictive validity was found for the PCL-R total score for all recidivism, but only for the 3-year follow-up period, and predictive accuracy for violence was low. A possible explanation is that female psychopaths do not rely on violent, intimidating behavior, but instead on more subtle, manipulative behavior (see Logan, Weizmann-Henelius, 2012). This is also in line with our finding that PCL-R Facet 1 Interpersonal was the only facet with significant predictive accuracy for all recidivism at 3-year follow-up, whereas in mainly male samples, PCL-R Facet Antisocial is usually the most predictive of recidivism (see Walters, Knight, Grann, & Dahle, 2008). Our results are

also in line with Richards et al. (2003) who found PCL-R Factor 1 to be most predictive in a sample of incarcerated female substance abusers.

The HCR-20^{V3} yielded more significant AUC values than the HCR-20. Several aspects of the HCR-20^{V3} are deemed better for violence risk assessment in women compared with the HCR-20 Version 2. For example, victimization, which is an important risk factor for females in both childhood and adulthood (Benda, 2005; van Voorhis et al., 2010), should now be considered for the entire lifespan instead of only during childhood. Furthermore, in HCR-20^{V3}, the psychiatric variables, Items H6 Major Mental Disorder and C3 Symptoms of Major Mental Disorder, are formulated broader and are better specified than in the HCR-20. Finally, the distinction in the HCR-20^{V3} between problematic behavior during different developmental stages and traumatic experiences is an important distinction for females.

The SAPROF and the START Strength scores did not demonstrate significant predictive validity in this female sample. In a previous study, the SAPROF showed better results for women (de Vries Robbé et al., 2016). However, in the referred study, the SAPROF was examined for inpatient violence and not for recidivism after discharge. It should also be noted that both tools were designed to predict short- to medium-term relapse, whereas in the present study, the overall follow-up period was substantially longer. Still, the START Vulnerability scores did predict significantly for both the medium- and long-term follow-up. Hence, the value of protective factors for recidivism after discharge for female forensic patients could not be confirmed in the current study.

Contrary to our hypothesis, the FAM total scores did not yield higher AUC values compared with the HCR-20^{V3} total scores for violent recidivism and not substantially higher compared with the HCR-20 total scores. In fact, the FAM performed slightly worse than the HCR-20^{V3}. Still, both the FAM total score and the final risk judgment nonviolent criminal behavior were significant predictors for all recidivism. The good predictive validity for the final risk judgment suggests that the process of coding risk assessment tools may help the assessor to deliberate more strongly about risks and thus provides support for the SPJ method. This is in line with a previous Dutch study (de Vogel & de Ruiter, 2005), although it should be noted that there was an overlap of 15 women in the samples. Nevertheless, the predictive accuracy of the FAM total score and the final risk judgment Violence for violent recidivism—for which it was primarily designed—was low. The FAM was originally developed in addition to HCR-20 Version 2. The findings from the current study suggest that the HCR-20^{V3} is improved with respect to the use with women and the FAM is less useful as an addition to HCR-20^{V3} compared with HCR-20. Possible explanations for the low predictive accuracy of the FAM for violent recidivism are that for some of the factors there is clear empirical support for the relation with general criminal offending, but not specifically for violence. Furthermore, for a number of items a correlation was found with previous violent behavior, but this does not necessarily mean that the factor is also related to future violent behavior. Still, mental health professionals have reported to consider the FAM a valuable addition to the HCR-20 and HCR-20^{V3}. Most importantly, they state that using the FAM—which does not take much extra time to code in addition to the HCR-20/HCR-20^{V3}—is valuable because it raises awareness of possible important gender issues in treatment and risk management (de Vogel & Louppen, 2017; Griswold et al., 2016). Furthermore, the FAM may be a valuable tool to predict risk outcomes other than convictions for violence after discharge, such as inpatient violence, self-destructive behavior, and victimization, which are also crucial to prevent.

LIMITATIONS

Several limitations to the present study should be acknowledged. First, the sample in this study was small and selective, as it only included forensic psychiatric patients who are characterized by severe offenses and complex psychopathology. Only discharged patients were included in the study and this might have caused a restricted range of scores on the instruments. It should also be noted that from the sample, nine women were readmitted to civil psychiatry, where reconviction is probably less likely to occur because of the institutionalization. Base rate of offending was rather low, especially for violent recidivism at the short follow-up period of 3 years ($n = 6$). Although ROC analyses are assumed to be insensitive to base rates, this low base rate and small sample could be lowering the results on predictive validity (see Flores et al., 2017; Mossman, 2013). Thus, the small and selective sample may have affected the results and may have resulted in an underestimation of reconvictions and predictive accuracy of the instruments. Still, the sample is representative for Dutch forensic psychiatry, as it included almost all of the discharged female forensic psychiatric patients in the Netherlands between 1993 and 2012. Considering the fact that females are a minority in forensic psychiatry, we believe that the current study adds to the scarce literature on (violence) risk assessment in women. However, the generalizability for more general gender-sensitive violence risk assessment is limited. Ideally, this study will be replicated in other samples, for instance, from the penitentiary system or general psychiatry.

Second, this study was retrospective, and although the files were extensive, not all items of the tools could be coded, especially not on the dynamic tool START, for which two cases had to be deleted for predictive validity analyses. For all other tools, the number of missing items was negligible. Prospective studies into the predictive validity of risk assessment tools in women would be valuable. However, considering females form a minority and have relatively low discharge rates and high mortality and readmission rates, this will take time. Another limitation is that we did not examine interrater reliability for all tools in the present study, but only for the HCR-20/FAM Historical items for a small sample from the total sample in this Dutch multicenter project into gender differences. Finally, an important limitation is that the outcome measure used in this study was only restricted to official reconviction data. The reconviction rate may be an underestimation of the actual recidivism because not all offenders are reported, apprehended, and arrested. Moreover, the outcome was only related to recidivism, while the FAM and START are also used to assess other types of risks, and both tools have previously been found to predict self-harm and inpatient violence (Campbell & Beech, 2018; O'Shea & Dickens, 2015). For future studies, it would be important to include other outcome measures, for instance, unofficial data on violence, data on victimization, and data on self-harm.

FUTURE RESEARCH

More research is definitely needed with multiple risk assessment tools—both SPJ tools and actuarial tools—into correctional and psychiatric samples of female offenders with several outcome measures, such as incidents during treatment, incidents of self-harm, victimization, or other negative outcomes. More specified studies, for example, into individual risk factors or more specific groups of female offenders, like sexual offenders, could yield valuable new insights. A direct comparison with a matched male sample is also

worthwhile to study. The present study comprised adult women. As it has been reported that the group of adolescent female offenders is growing, there is an urgent need to conduct more studies into the value of risk assessment and management in girls and young women (see also Emeka & Sorensen, 2009; Moretti et al., 2005). A recent meta-analysis found good results for an actuarial tool, the Youth Level of Service (YLS) for both male and female juvenile offenders (Pusch & Holtfreter, 2017). This corresponds with the review by Geraghty and Woodhams (2015) who found the strongest results for the adult version, the LSI (see also the results of the meta-analysis of Olver et al., 2014). More research applying both SPJ instruments and actuarial tools, especially those that include dynamic risk factors, would be valuable. In the present study, we only examined SPJ risk assessment tools—next to the PCL-R—as these are mandated to use in Dutch forensic psychiatry. Finally, considering the results on particularly PCL-R Facet 1 Interpersonal, we strongly recommend to conduct more studies into the different psychopathic features in women.


CLINICAL IMPLICATIONS

Overall, this study shows that SPJ tools such as the START, the HCR-20^{V3}, and the FAM as well as the PCL-R have some, but not very convincing, predictive value in adult female forensic psychiatric populations. The assessor should thus exert caution in the interpretation of the results of risk assessments, especially when important decisions need to be made, for example, regarding mandatory admittance to a hospital. We strongly recommend more research, but based on the results of the present study, we cautiously recommend to use the HCR-20^{V3} for violence risk assessment in forensic mental health care and the START for more general risk assessment, especially when short- to medium-term assessments are needed or when only recent information to code dynamic risk factors is available. The FAM may be a useful addition to the HCR-20^{V3} for more gender-sensitive risk assessment and management, mainly for clinical purposes but not for improving predictive power. Furthermore, for clinical practice, it is important to not only consider risk of violence or general recidivism in society but also pay attention to the other risks, such as self-destructive behavior, victimization, or substance use.

NOTE

1. It should be noted that 15 of these 78 women have been examined in a previous study (de Vogel & de Ruiter, 2005), however only with the Psychopathy Checklist-Revised (PCL-R) and HCR-20 and with a shorter follow-up period. For the present study, the Historical, Clinical, Risk management-20 (HCR-20) was coded again and new reconviction data were retrieved.

ORCID ID

Vivienne de Vogel  <https://orcid.org/0000-0001-7671-1675>

REFERENCES

- Adams, S. R., & Freeman, D. R. (2002). Women who are violent: Attitudes and beliefs of professionals working in the field of domestic violence. *Military Medicine, 167*, 445-450. doi:10.1093/milmed/167.6.445
- Andrews, D. A., & Bonta, J. (2000). *The level of service-inventory-revised*. Toronto, Ontario, Canada: Multi-Health Systems.
- Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, L. J., Brews, A., & Wormith, J. S. (2012). Are the major risk/need factors predictive of both female and male reoffending? A test with the eight domains of the level of service/case management inventory. *International Journal of Offender Therapy and Comparative Criminology, 56*, 113-133. doi:10.1177/0306624X10395716

- Benda, B. (2005). Gender differences in life-course theory of recidivism: A survival analysis. *International Journal of Offender Therapy and Comparative Criminology*, 49, 325-342. doi:10.1177/0306624X04271194
- Binswanger, I. A., Stern, M. F., Deyo, R. A., Heagerty, P. J., Cheadle, A., Elmore, J. G., & Koepsell, T. D. (2007). Release from prison—A high risk of death for former inmates. *New England Journal of Medicine*, 356, 157-165. doi:10.1056/NEJMsa064115
- Blanchette, K., & Brown, S. L. (2006). *Wiley Series in Forensic Clinical Psychology: The assessment and treatment of women offenders: An integrative perspective*. Chichester, UK: John Wiley.
- 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 and Behavior*, 39, 1481-1508. doi:10.1177/0093854812456777
- Brown, D. W., Anda, R. F., Tiemeier, H., Felitti, V. J., Edwards, V. J., Croft, J. B., & Giles, W. H. (2009). Adverse childhood experiences and the risk of premature mortality. *American Journal of Preventive Medicine*, 37, 389-396. doi:10.1016/j.amepre.2009.06.021
- Burman, M. J., Batchelor, S. A., & Brown, J. A. (2001). Researching girls and violence: Facing the dilemmas of fieldwork. *British Journal of Criminology*, 41, 443-459. doi:10.1093/bjc/41.3.443
- 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, 914-933. doi:10.1080/14789949.2018.1477975
- Chesney, E., Goodwin, G. M., & Fazel, S. (2014). Risks of all-cause and suicide mortality in mental disorders: A meta-review. *World Psychiatry*, 13, 153-160. doi:10.1002/wps.20128
- Coid, J., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Roberts, C., . . . Rogers, R. D. (2009). Gender differences in structured risk assessment: Comparing the accuracy of five instruments. *Journal of Consulting and Clinical Psychology*, 77, 337-348. doi:10.1037/a0015155
- Davidson, J. T., & Chesney-Lind, M. (2009). Discounting women: Context matters in risk and need assessment. *Critical Criminology*, 17, 221-245. doi:10.1007/s10612-009-9084-x
- 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, 70-74. doi:10.1192/bjp.bp.106.029215
- Desmarais, S. L., Reeves, K. A., Nicholls, T. L., Telford, R. P., & Fiebert, M. S. (2012). Prevalence of physical violence in intimate relationships, Part 2: Rates of male and female perpetration. *Partner Abuse*, 3, 170-198. doi:10.1891/1946-6560.3.2.170
- de Vogel, V. (2005). *Structured risk assessment of (sexual) violence in forensic clinical practice: The HCR-20 and SVR-20 in Dutch forensic psychiatric patients*. Amsterdam, The Netherlands: Dutch University Press. Retrieved from <http://dare.uva.nl/document/13725>
- de Vogel, V., & de Ruiter, C. (2005). The HCR-20 in personality disordered female offenders: A comparison with a matched sample of males. *Clinical Psychology & Psychotherapy*, 12, 226-240. doi:10.1002/cpp.452
- de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbé, M. (2012). *SAPROF: Guidelines for the assessment of protective factors for violence risk* (2nd ed.). Utrecht, The Netherlands: Forum Educatief.
- de Vogel, V., de Vries Robbé, M., van Kalmthout, W., & Place, C. (2014). *Female Additional Manual (FAM): Additional guidelines to the HCR-20V3 for assessing risk for violence in women*. Utrecht, The Netherlands: Van der Hoeven Kliniek. Available from www.violencebywomen.com
- de Vogel, V., & Lancel, M. (2016). Gender differences in the manifestation of psychopathy: Results from a multicentre study in forensic psychiatry. *International Journal of Forensic Mental Health*, 15, 97-110. doi:10.1080/14999013.2016.1138173
- de Vogel, V., & Louppen, M. (2017). Female offenders. In R. Roesch & A. N. Cook (Eds.), *Handbook of forensic mental health services* (pp. 369-400). London, England: Routledge.
- de Vogel, V., & Nicholls, T. L. (2016). Gender matters: An introduction to the special issue on women and girls. *International Journal of Forensic Mental Health*, 15, 1-25. doi:10.1080/14999013.2016.1141439
- de Vogel, V., Stam, J., Bouman, Y., Ter Horst, P., & Lancel, M. (2016). Violent women: A multicentre study into gender differences in forensic psychiatric patients. *The Journal of Forensic Psychiatry & Psychology*, 27, 145-168. doi:10.1080/14789949.2015.1102312
- de Vogel, V., van den Broek, E., & de Vries Robbé, M. (2014). The value of the HCR-20V3 in Dutch forensic clinical practice. *International Journal of Forensic Mental Health*, 13, 109-121. doi:10.1080/14999013.2014.906518
- de Vries Robbé, M., de Vogel, V., Wever, E., Douglas, K. S., & Nijman, H. L. I. (2016). Risk and protective factors for inpatient aggression. *Criminal Justice and Behavior*, 43, 1364-1385. doi:10.1177/0093854816637889
- Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20V3: Assessing risk of violence—User guide*. Burnaby, British Columbia, Canada: Simon Fraser University.
- Douglas, K. S., Shaffer, C., Blanchard, A. J. E., Guy, L. S., Reeves, K., & Weir, J. (2017). HCR-20 violence risk assessment scheme: Overview and annotated bibliography. Retrieved from <https://kdouglas.files.wordpress.com/2017/06/hcr-20-annotated-bibliography-version-13.pdf>

- Emeka, T. Q., & Sorensen, J. R. (2009). Female juvenile risk: Is there a need for gendered assessment instruments? *Youth Violence and Juvenile Justice*, 7, 313-330. doi:10.1177/1541204009334083
- Flores, A. W., Holsinger, A. M., Lowenkamp, C. T., & Cohen, T. H. (2017). Time-free effects in predicting recidivism using both fixed and variable follow-up periods: Do different methods produce different results. *Criminal Justice and Behavior*, 44, 121-137. doi:10.1177/0093854816678649
- Forouzan, E., & Cooke, D. J. (2005). Figuring out la femme fatale: Conceptual and assessment issues concerning psychopathy in females. *Behavioral Sciences & the Law*, 23, 765-778. doi:10.1002/bsl.669
- Garcia-Mansilla, A., Rosenfeld, B., & Nicholls, T. L. (2009). Risk assessment: Are current methods applicable to women? *International Journal of Forensic Mental Health*, 8, 50-61. doi:10.1080/14999010903014747
- Geraghty, K. A., & Woodhams, J. (2015). The predictive validity of risk assessment tools for female offenders: A systematic review. *Aggression and Violent Behavior*, 21, 25-38. doi:10.1016/j.avb.2015.01.002
- 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, 48-64. doi:10.1080/14999013.2015.1134726
- Greig, D. (2014). *Gender-sensitive violence risk assessment: A preliminary investigation into the predictive accuracy and clinical utility of the Female Additional Manual*. Burnaby, British Columbia, Canada: Simon Fraser University.
- Griswold, H., Green, D., Belfi, B., Grossi, L., Smith, J., & Otten, J. (2016, June). *The Female Additional Manual (FAM): An investigation of its predictive validity among female defendants adjudicated not guilty by reason of insanity*. Paper Presented at the Annual Conference of the International Association for Forensic Mental Health Services, New York, NY.
- Hare, R. D. (2003). *Manual for the hare psychopathy checklist-revised*. Toronto, Ontario, Canada: Multi-Health Systems.
- Hart, J. L., O'Toole, S. L., Price-Sharps, J. L., & Shaffer, T. W. (2007). The risk and protective factors of violent juvenile offending: An examination of gender differences. *Youth Violence and Juvenile Justice*, 5, 367-384. doi:10.1177/1541204006297367
- Hildebrand, M., de Ruiter, C., de Vogel, V., & van der Wolf, P. (2002). Reliability and factor structure of the Dutch language version of Hare's psychopathy checklist-revised. *International Journal of Forensic Mental Health*, 1, 139-154. doi:10.1080/14999013.2002.10471169
- Kim, H. K., Capaldi, D. M., Pears, K. C., Kerr, D. C. R., & Owen, L. D. (2009). Intergenerational transmission of internalising and externalising behaviours across three generations: Gender-specific pathways. *Criminal Behaviour and Mental Health*, 19, 125-141. doi:10.1002/cbm.708
- Leistico, A. R., Salekin, R. T., DeCoster, J., & Rogers, R. (2008). A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. *Law and Human Behavior*, 32, 28-45. doi:10.1007/s10979-007-9096-6
- Levene, K. S., Augimeri, L. K., Pepler, D. J., Walsh, M. M., Webster, C. D., & Koegl, C. J. (2001). *Early Assessment Risk List for Girls: EARL-21G (Version 1—Consultation version)*. Toronto, Ontario, Canada: Earls court Child and Family Centre.
- Logan, C., & Weizmann-Henelius, G. (2012). Psychopathy in women: Presentation, assessment, and management. In H. Häkkinen-Nyholm & J. Nyholm (Eds.), *Psychopathy and law: A practitioner's guide* (pp. 99-125). Singapore: Wiley-Blackwell.
- May-Chahal, C., & Cawson, P. (2005). Measuring child maltreatment in the United Kingdom: A study of the prevalence of child abuse and neglect. *Child Abuse & Neglect*, 29, 969-984. doi:10.1016/j.chiabu.2004.05.009
- McKeown, A. (2010). Female offenders: Assessment of risk in forensic settings. *Aggression and Violent Behavior*, 15, 422-429. doi:10.1016/j.avb.2010.07.004
- Messer, J., Maughan, B., Quinton, D., & Taylor, A. (2004). Precursors and correlates of criminal behaviour in women. *Criminal Behaviour and Mental Health*, 14, 82-107. doi:10.1002/cbm.575
- Miller, S., Malone, P. S., & Dodge, K. A. (2010). Developmental trajectories of boys' and girls' delinquency: Sex differences and links to later adolescent outcomes. *Journal of Abnormal Child Psychology*, 38, 1021-1032. doi:10.1007/s10802-010-9430-1
- Moretti, M. M., Catchpole, R. E. H., & Odgers, C. (2005). The dark side of girlhood: Recent trends, risk factors and trajectories to aggression and violence. *The Canadian Child and Adolescent Psychiatry Review*, 14, 21-25.
- Morgan, M., & Patton, P. (2002). Gender-responsive programming in the justice system—Oregon's guidelines for effective programming for girls. *Federal Probation*, 66, 57-65.
- Mossman, D. (2013). Evaluating risk assessments using receiver operating characteristic analysis: Rationale, advantages, insights, and limitations. *Behavioral Sciences & the Law*, 31, 23-39. doi:10.1002/bsl.2050
- Nicholls, T. L., Cruise, K. R., Greig, D., & Hinz, H. (2015). Female offenders. In B. L. Cutler & P. A. Zapf (Eds.), *APA handbook of forensic psychology: Criminal investigation, adjudication, and sentencing outcomes* (Vol. 2, pp. 79-123). Washington, DC: American Psychological Association. doi:10.1037/14462-004
- Nicholls, T. L., Ogloff, J. P., Brink, J., & Spidel, A. (2005). Psychopathy in women: A review of its clinical usefulness for assessing risk for aggression and criminality. *Behavioral Sciences & the Law*, 23, 779-802. doi:10.1002/bsl.678

- Ogders, C. L., Moretti, M. M., & Reppucci, N. D. (2005). Examining the science and practice of violence risk assessment with female adolescents. *Law and Human Behavior, 29*, 7-27. doi:10.1007/s10979-005-1397-z
- Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2014). Thirty years of research on the Level of Service Scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment, 26*, 156-176. doi:10.1037/a0035080
- O'Shea, L. E., & Dickens, G. L. (2015). Predictive validity of the Short-Term Assessment of Risk and Treatability (START) for aggression and self-harm in a secure mental health service: Gender differences. *International Journal of Forensic Mental Health, 14*, 132-146. doi:10.1080/14999013.2015.1033112
- 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 and Violent Behavior, 18*, 255-270. doi:10.1016/j.avb.2012.11.016
- Otto, R. K., & Douglas, K. S. (Eds.). (2010). *International Perspectives on Forensic Mental Health: Handbook of violence risk assessment*. New York, NY: Routledge.
- Pusch, N., & Holtfreter, K. (2017). Gender and risk assessment in juvenile offenders: A meta-analysis. *Criminal Justice and Behavior, 45*, 56-81. doi:10.1177/0093854817721720
- Rettinger, L. J., & Andrews, D. A. (2010). General risk and need, gender specificity, and the recidivism of female offenders. *Criminal Justice and Behavior, 37*, 29-46. doi:10.1177/0093854809349438
- 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. doi:10.1007/s10979-005-6832-7
- Richards, H. J., Casey, J. O., & Lucente, S. W. (2003). Psychopathy and treatment response in incarcerated female substance abusers. *Criminal Justice and Behavior, 30*, 251-276. doi:10.1177/0093854802251010
- Rodermond, E., Kruttschnitt, C., Slotboom, A. M., & Bijleveld, C. C. (2016). Female desistance: A review of the literature. *European Journal of Criminology, 13*, 3-28. doi:10.1177/1477370815597251
- Sahota, S., Davies, S., Duggan, C., Clarke, M., Huband, M., & Owen, V. (2010). Women admitted to medium secure care: Their admission characteristics and outcome as compared with men. *International Journal of Forensic Mental Health, 9*, 110-117. doi:10.1080/14999013.2010.499555
- Salisbury, E. J., Boppre, B., & Kelly, B. (2016). Implications for the treatment of justice-involved women. In F. S. Taxman (Ed.), *Handbook on risk and need assessment: Theory and practice* (pp. 220-243). New York, NY: Routledge.
- Skeem, J., Schubert, C., Stowman, S., Beeson, S., Mulvey, E., Gardner, W., & Lidz, C. (2005). Gender and risk assessment accuracy: Underestimating women's violence potential. *Law and Human Behavior, 29*, 173-186. doi:10.1007/s10979-005-3401-z
- van Voorhis, P., Bauman, A., & Brushett, R. (2013). *Revalidation of the women's risk needs assessment: Institutional results—Final report*. Cincinnati, OH: Center for Criminal Justice Research.
- van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior, 37*, 261-288. doi:10.1177/0093854809357442
- Vaurio, O., Repo-Tiihonen, E., Kautiainen, H., & Tiihonen, J. (2018). Psychopathy and mortality. *Journal of Forensic Sciences, 63*, 474-477. doi:10.1111/1556-4029.13566
- Viljoen, S., Nicholls, T. L., Greaves, C., de Ruiter, C., & Brink, J. (2011). Resilience and successful community reintegration among female forensic psychiatric patients: A preliminary investigation. *Behavioral Sciences & the Law, 29*, 752-770. doi:10.1002/bsl.1001
- Viljoen, S., Nicholls, T. L., Roesch, R., Gagnon, N., Douglas, K. S., & Brink, J. (2016). Exploring gender differences in the utility of strength based risk assessment measures. *International Journal of Forensic Mental Health, 15*, 149-163. doi:10.1080/14999013.2016.1170739
- Völlm, B. A., & Dolan, M. C. (2009). Self-harm among UK prisoners: A cross-sectional study. *Journal of Forensic Psychiatry & Psychology, 20*, 741-751. doi:10.1080/14789940903174030
- Walmsley, R. (2015). *World female imprisonment list: Women and girls in penal institutions, including pre-trial detainees /remand prisoners* (3rd ed.). Retrieved from www.prisonstudies.org/sites/default/files/resources/downloads/world_female_imprisonment_list_third_edition_0.pdf
- Walters, G. D., Knight, R. A., Grann, M., & Dahle, K. P. (2008). Incremental validity of the psychopathy checklist facet scores: Predicting release outcome in six samples. *Journal of Abnormal Psychology, 117*, 396-405. doi:10.1037/0021-843X.117.2.396
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing the risk of violence* (Version 2). Burnaby, British Columbia, Canada: Simon Fraser University.
- Webster, C. D., Martin, M., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *Short-Term Assessment of Risk and Treatability (START): Clinical guide for evaluation risk and recovery* (Version 1.1). Hamilton, Ontario, Canada: St. Joseph's Healthcare Hamilton.
- Yesberg, J. A., Scanlan, J. M., Hanby, L. J., Serin, R. C., & Polaschek, D. L. L. (2015). Predicting women's recidivism: Validating a dynamic community-based "gender-neutral" tool. *Probation Journal, 62*, 33-48. doi:10.1177/0264550514562851

Vivienne de Vogel, PhD, is a psychologist and professor at the HU University of Applied Sciences Utrecht, Center for Social Innovation. She also works at the Department of Research at the De Forensische Zorgspecialisten (a forensic center) in Utrecht, The Netherlands. Her research focuses on gender differences in forensic care, violence risk assessment and management, psychopathy, and inpatient violence.

Mieke Bruggeman, MSc, is a criminologist working as a junior researcher at the Netherlands Institute for the Study of Crime and Law Enforcement in Amsterdam, The Netherlands. Her research focuses on violence in closed youth care institutions and intellectual disabled young people and their place in the criminal procedure and in care.

Marika Lancel, PhD, is a psychologist and somnologist. She is head of the research department of the forensic hospital in Assen, The Netherlands, and professor at the University of Groningen, The Netherlands. The main theme of her research is the influence of sleep disorders on psychopathology, such as impulsivity and aggression.