

Assessment *The HCR-20 in Personality Disordered Female Offenders: A Comparison with a Matched Sample of Males*

Vivienne de Vogel,^{1*} and Corine de Ruiter^{2,3}

¹*Dr. Henri van der Hoeven Kliniek, Utrecht, The Netherlands*

²*University of Amsterdam, Amsterdam, The Netherlands*

³*Trimbos Institute (Netherlands Institute of Mental Health and Addiction), Utrecht, The Netherlands*

This study examines the predictive validity of the HCR-20 in a sample of 42 female patients admitted to a Dutch forensic psychiatric hospital. The findings are compared with those for a matched sample of 42 male forensic psychiatric patients. The inter-rater reliability of the HCR-20 was good for both female and male patients. There were significant differences between female and male patients in mean HCR-20 item scores, but the mean H, C and R-subscale scores and total score were comparable. The base rate for inpatient violence was similar for female (30%) and male patients (29%), but the base rate for violent recidivism after discharge was significantly higher for the male sample (43%) compared with the female sample (13%). For male patients, the HCR-20 demonstrated good to excellent predictive validity for violent outcome (violent recidivism and inpatient violence); however, predictive accuracy for female patients was much lower. In females, only the HCR-20 final risk judgment, and not the HCR-20 total score, demonstrated significant predictive validity for violent outcome. Copyright © 2005 John Wiley & Sons, Ltd.

Gender is one of the most significant predictors of violence; regardless of age, ethnicity, culture and socioeconomic status, men are significantly more often convicted for violent offenses than women (Archer & McDaniel, 1995; Monahan et al., 2001). However, research also suggests that mental disorder reduces the gender gap in violence, especially for inpatient aggression. Among psychiatric

patients, the base rate for (inpatient) violence is not significantly different for male and female patients (Lidz, Mulvey, & Gardner, 1993; McNiel & Binder, 1990; Newhill, Mulvey, & Lidz, 1995; Nicholls, unpublished matter's thesis; Tardiff, Marzuk, Leon, Portera, & Wiener, 1997). Ross, Hart and Webster (1998) found no sex differences between a sample of 82 male and 49 female psychiatric patients in the occurrence of inpatient aggression. However, regarding violence in the community after treatment, male patients were found to be four times more likely than female patients to express any aggression.

*Correspondence to: Vivienne de Vogel, M.A., Department of Research, Dr. Henri van der Hoeven Kliniek, P.O. Box 174, 3500 AD Utrecht, The Netherlands.
Email: vdevogel@hoevenstichting.nl

Research has demonstrated that unstructured clinical judgment of violence risk is sensitive to sex-based biases; clinicians tend to underestimate the risk of violence in female psychiatric patients (Lidz et al., 1993; McNeil & Binder, 1995). More generally, research has revealed some important limitations of unstructured clinical judgment, such as poor reliability and validity (see Monahan, 1981; Quinsey, Harris, Rice, & Cormier, 1998). Use of structured risk assessment instruments is recommended to avoid these types of bias and to optimize the reliability and validity of violence risk assessment (Borum, 1996). A problem, however, is that existing structured risk assessment instruments are developed based on violence risk research primarily in male samples. Thus, the question arises of whether the risk factors for violence found in male samples are also valid for females and, consequently, whether the existing structured risk assessment instruments are suitable for use with female patients. Several authors have argued that risk factors for violence in female samples are generally the same as in male samples and that existing risk assessment instruments are likely valid for use with females (Blanchette, 1997; Harer & Langan, 2001; Simourd & Andrews, 1994; Strand & Belfrage, 2001). Loucks and Zamble (1999) compared the characteristics of 100 female offenders to a sample of male offenders,¹ and although they found some differences in the occurrence of important life experiences these differences were not predictive of criminal behavior. In contrast, others have argued that assessing risk for violence is different for women compared with men because risk factors for women are closely linked to their unique experiences as women, for instance victimization (Chesney-Lind, 1989; Scarth & McLean, 1994) or to the fact that social bonds are of greater importance to women than to men and that women are thus more sensitive to disruptions in close relationships (see Funk, 1999; Odgers & Moretti, 2002). Funk (1999) tested risk factors for reoffending in 388 male and 112 female juvenile delinquents on probation and found several risk factors (e.g. child abuse or neglect, running away from home) that were significantly predictive for females but not for males. Therefore, she concluded that risk factors for females differ substantially from those of their male counterparts, that risk assessment instruments fail to identify most

female risk factors, and that separate risk assessment instruments for males and females should improve classifications for risk of reoffending. To our knowledge, only one structured risk assessment instrument has been developed especially for the assessment of risk in females: the Early Assessment Risk List for Girls (EARL-21G; Levene et al., 2001). Vitale and Newman (2001) stated that existing risk assessment instruments have not yet been adequately tested to determine their generalizability to women.

A structured risk assessment instrument that has drawn considerable international attention is the Historical, Clinical, Risk Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997). The HCR-20 is a checklist according to the structured professional judgment (SPJ) approach. In the SPJ approach, the risk assessment is performed by a forensic clinician by means of a standardized checklist, containing empirically derived risk factors for violence, historical (fixed) as well as dynamic. The HCR-20 consists of 20 items representing risk factors for violence in the past (Historical scale), present (Clinical scale) and future (Risk management scale). Research with various psychiatric and forensic samples in different countries has demonstrated good inter-rater reliability and predictive validity for the HCR-20 (see Douglas & Weir, 2003). The HCR-20 was primarily developed on the basis of research in male samples and most research into the psychometric properties of the HCR-20 has been done in male samples. Therefore, the question of whether the HCR-20 is also suitable for use with females seems important. Nicholls (unpublished master's thesis) conducted a retrospective study to evaluate the validity of the HCR-20 and the Psychopathy Checklist: Screening Version (PCL:SV; Hart, Cox, & Hare, 1995) for assessing female patients' risk for inpatient and community violence. She compared the results of 47 female patients with a matched sample of 47 male patients admitted to a forensic psychiatric hospital and found the distribution of the mean HCR-20 and PCL:SV scores to be comparable. The HCR-20 showed good predictive accuracy for inpatient aggression for both male and female patients. The predictive accuracy of the HCR-20 for community aggression was modest for both samples. Strand and Belfrage (2001) conducted a retrospective study to investigate the utility of the HCR-20 in a female forensic psychiatric sample. They compared the HCR-20 scores of 63 female and 85 male patients admitted to two forensic psychiatric hospitals in Sweden and found

¹The authors do not mention the number of males or whether the males were matched to the females.

some significant differences in mean individual item scores, however, the mean subscale scores and total score did not differ significantly. The authors thus concluded that the HCR-20 is suitable for use in female forensic psychiatric patients, particularly to assess inpatient violence. A limitation of this research is that the authors did not examine the predictive validity of the HCR-20 scores for violent outcome, which is the most important aspect to decide whether the HCR-20 is adequate for female patients.

An important issue to keep in mind when assessing risk for future violence is that violence is a multifaceted construct. Risk assessment should not only be directed at predicting the likelihood of violence, but also take into account the severity, nature, frequency and imminence of violence (Hart, 1998). Research has shown that in general, the nature, severity and victims of violent offenses committed by women are different from those committed by men. Female violence is less often sexual in nature, less often characterized as instrumental and more often as reactive, less often resulting in injury, more often relational and more often occurring in the residence (Monahan et al., 2001; Nicholls, unpublished master's thesis; Odgers & Moretti, 2002). Summarizing the above suggests that the factors and assessment of violence risk differ at least to a certain extent between female and male patients, and that the utility of the existing structured risk assessment instruments for women has yet to be convincingly proven.

In this article, we will present findings on the inter-rater reliability and predictive validity of the HCR-20 in a sample of 42 female patients who have been admitted to the Dr. Henri van der Hoeven Kliniek, a Dutch forensic psychiatric hospital. The findings are compared with those for a matched sample of 42 male forensic psychiatric patients from the same hospital. The aim of the present study was to examine whether there are differences between female and male forensic psychiatric patients regarding mean HCR-20 scores, inter-rater reliability and predictive validity for violent outcome. In addition, we coded the Psychopathy Checklist-Revised (PCL-R; Hare, 1991) and compared the mean scores and predictive validity for violent outcome between female and male patients. Several studies have been conducted into the use of the PCL-R in female samples. In general, a lower prevalence of psychopathy among females compared with males was found (Grann, 2000; Salekin, Rogers, & Sewell, 1997; Vitale, Smith,

Brinkley, & Newman, 2002; Warren et al., 2003). Vitale and Newman (2001) reviewed the literature regarding the PCL-R in female samples and found good support for the reliability, but modest support for the predictive validity. They concluded that whereas the PCL-R might be able to postdict violent behavior in the past, there is no evidence that the PCL-R can predict future violence in women. The issue of whether the PCL-R is suitable for the assessment of psychopathy in women is not settled. Some have argued that the PCL-R is adequate for assessing psychopathy in women, since they found a considerable degree of similarity to the construct of psychopathy in male offenders (Salekin et al., 1997; Warren et al., 2003). In contrast, Vitale and colleagues (2002) believe that the findings thus far are not sufficiently convincing to conclude a similarity of the PCL-R structure across gender. They express concern that some PCL-R items do not adequately assess the construct of psychopathy as it is expressed in women.

METHOD

Setting

This study was conducted at the Dr. Henri van der Hoeven Kliniek, a forensic psychiatric hospital in The Netherlands with 135 patients. Patients are admitted under the judicial measure *terbeschikkingstelling* (tbs), which is translated as 'disposal to be treated on behalf of the state'. The tbs-order is imposed by court on offenders who have committed a serious offense and are considered to have diminished responsibility for it because of severe psychiatric disorder. The tbs-order is of indefinite duration; every one or two years the court re-evaluates the patient to determine whether the risk of recidivism is still too high and treatment needs to be continued.

The Dr. Henri van der Hoeven Kliniek was founded in 1955 and is one of 13 forensic psychiatric institutions in The Netherlands. The hospital provides a variety of treatment activities, for instance job training, education, sports, creative arts, and psychotherapy. The treatment model of the hospital is cognitive-behavioral with an emphasis on relapse prevention, in which the 'no cure but control' principle dominates (Laws, Hudson, & Ward, 2000). The hospital is one of three forensic psychiatric hospitals in The Netherlands that admit both male and female offenders under the tbs-order. The proportion of female patients in

the hospital is about 15%. Female patients do not stay on a separate ward but reside among the men in living-groups, although there should be at least two women in one living-group. There are specific treatment activities for female patients, such as female sports and a therapy group that meets weekly . . . In this therapy group, there are several modules about themes relevant to female patients, for instance, what it is like to live in a predominantly male environment, victimization, and sexuality.

Procedure

First, we collected archival data from the hospital records for 42 female patients admitted to the hospital between 1985 and 2003. The Dutch versions of the HCR-20 (Philipse, de Ruiter, Hildebrand, & Bouman, 2000) and PCL-R (Vertommen, Verheul, de Ruiter, & Hildebrand, 2002) were coded for all 42 women on the basis of all available file information. There were three categories. (1) 15 women whose HCR-20 had already been coded in a recently conducted retrospective study into the predictive validity of the HCR-20 (see de Vogel, de Ruiter, Hildebrand, Bos, & van de Ven, 2004). In this study, the rating procedure was performed while all raters were blind to reconviction data. (2) 23 females from an ongoing prospective study in which the HCR-20 is coded independently by a researcher, a treatment supervisor and a group-leader. During a case conference, the raters discuss their scores and agree upon a consensus score that was used for the analyses in the present study (see de Vogel & de Ruiter, 2004). (3) Four female patients who were admitted to the hospital at the time of the current study, but had not been included in the prospective study mentioned above. For these four cases, two raters independently and prospectively coded the HCR-20 and agreed upon a consensus score that was used in the analyses. In order to establish the inter-rater reliability, we used all codings performed by three independent raters, i.e. the 23 codings from the prospective study and four cases from the retrospective study. The mean follow-up period of female patients from the retrospective study was 74.6 months (SD = 23.9, range = 26.7–109.6) and that from the prospective study 10.2 months (SD = 7.8, range = 0.2–26.3).

Second, we matched the women to 42 male patients on year of birth, type of index offense, ethnicity, and type of psychopathology (i.e. Axis I,

Axis II or comorbid Axis I and II, according to the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders*, DSM-IV; APA, 1994). Regarding the index offenses, there were two women with a property offense without violence, and only one male with the same index offense, so we decided to match one of these two women to a male with a property offense in combination with violence (see Table 1). The 42 men were identified from a total sample of 205 male patients admitted between 1985 and 2003 and obtained from two sources: (1) 21 cases from the recently conducted retrospective study into the predictive validity of the HCR-20 (see above); (2) 21 cases from the ongoing prospective study (see above). Inter-rater reliability was established for all codings performed by three independent raters, i.e. 21 cases from the prospective study and seven of the 21 cases from the retrospective study. For male patients from the retrospective study the mean follow-up period was 81.1 months (SD = 23.8, range = 46.1–114.9), and for the males from the prospective study it was 18.7 months (SD = 6.6, range = 4.7–26.3). The mean follow-up period of men from the prospective study was significantly longer than the mean follow-up period of women from the prospective study ($F = 1.2, p < 0.01$). For the retrospective study, the mean follow-up period of men and women did not differ significantly ($F = 0.53, p = 0.42$).

Subjects

Table 1 presents demographic, psychiatric and criminal characteristics for the female and male samples. There were a number of significant differences between female and male patients. Female patients compared with male patients were more often involved in an intimate relationship at the time of the index offense, had less often abused substances, were diagnosed more often with borderline personality disorder (BPD) and less often with narcissistic personality disorder (NPD), obtained higher scores on intelligence scales, particularly on verbal intelligence, and were older at the time of their first conviction. Antisocial personality disorder (ASPD) was less prevalent in women than in men, although the difference was marginally significant ($\chi^2 = 3.6, p = 0.06$). Only the antisocial, borderline and narcissistic personality disorders (according to the DSM-IV) are reported in Table 1 because these disorders are the most prevalent in both men and women in forensic

Table 1. Sample characteristics

	Female patients N = 42	Male patients N = 42
Demographic		
Mean age upon admission	33.2	30.7
Dutch nationality	38 (91%)	40 (95%)
Upbringing in foster or children's home	13 (31%)	17 (41%)
Single (at the time of the index offense)	24 (57%)	36 (86%)**
No education after primary school	19 (45%)	21 (50%)
Unemployed (at the time of the index offense)	38 (91%)	33 (79%)
Psychiatric		
Prior out-patient treatment(s)	22 (52%)	15 (36%)
Prior in-patient admission(s)	23 (55%)	24 (57%)
Substance abuse	27 (64%)	35 (83%)*
Antisocial personality disorder [†]	8 (25%)	14 (48%)
Borderline personality disorder [†]	24 (75%)	7 (24%)**
Narcissistic personality disorder [†]	3 (9%)	10 (35%)*
Mean intelligence scores: total ^{††}	111.3	105.1
Mean intelligence scores: verbal ^{††}	115.4	97.4**
Mean intelligence scores: performance ^{††}	114.8	111.8
Offenses		
(Attempted) Homicide	26 (62%)	26 (62%)
Sexual	1 (2%)	1 (2%)
Violent	4 (10%)	5 (12%)
Arson	9 (21%)	9 (21%)
Property	2 (5%)	1 (2%)
Victim was not a stranger	31 (74%)	26 (62%)
Victim was (ex-) partner or relative	15 (36%)	8 (19%)
Mean duration of imprisonment in months	19.5	28.1
Mean number of previous convictions	1.9	6.4
Mean age at first conviction	27.2	21.1**

** $p < 0.01$. * $p < 0.05$ (two tailed).

[†] Personality disorders were diagnosed with the SIDP-IV (Pfohl, Blum, & Zimmerman, 1995) and available for 32 females and 29 males.

^{††} Mean intelligence scores were available for 18 females and 21 males.

psychiatric settings (see Coid, Kahtan, Gault, & Jarman, 1999; Hildebrand & de Ruiter, 2004; de Ruiter & Greeven, 2000; Warren et al., 2002). There was a trend that women more often had a relative or (ex-) partner as victim ($\chi^2 = 2.9$, $p = 0.08$).

Instruments

HCR-20

The HCR-20 is a structured professional guideline (checklist) designed for the assessment of risk of future violence in adult offenders with a violent history and/or a major mental disorder or personality disorder. The instrument was developed from a thorough consideration of the empirical literature and the clinical expertise of a number of forensic clinicians. The HCR-20 consists of 20 items,

divided into three subscales, Historical scale, Clinical scale and Risk management scale, that relate to risk factors in the past, present and future, respectively (see Table 2 below). The Historical items are static, unchangeable factors,² whilst the Clinical and Risk management factors are considered to be changeable, for instance due to clinical intervention. The items have to be coded on a three point scale: '0', item does not apply according to the available information '1', the item probably or partially applies, and '2', the item definitely applies. Aside from the 20 items, the HCR-20 offers the possibility to code 'other considerations', that is, case-

²This is not completely true; Historical items can change in an unfavorable direction. For instance, the score on item 10 increases when a patient violates the rules by escaping from the secure hospital.

Table 2. Mean HCR-20 and PCL-R scores (standard deviations in brackets), final risk judgments and psychopathy diagnosis

	Female patients N = 42	Male patients N = 42
Historical items		
1. Previous violence	2.0 (0.22)	1.9 (0.26)
2. Young age at first violent incident	1.2 (0.65)	1.5 (0.55)*
3. Relationship instability	1.9 (0.30)	1.7 (0.51)*
4. Employment problems	1.4 (0.80)	1.5 (0.67)
5. Substance use problems	1.3 (0.87)	1.5 (0.80)
6. Major mental illness	0.83 (0.85)	0.90 (0.85)
7. Psychopathy	0.38 (0.54)	0.71 (0.71)*
8. Early maladjustment	1.7 (0.52)	1.8 (0.40)
9. Personality disorder	1.9 (0.45)	1.9 (0.33)
10. Prior supervision failure	1.4 (0.80)	1.4 (0.86)
Clinical items		
1. Lack of insight	1.4 (0.54)	1.5 (0.63)
2. Negative attitudes	0.98 (0.78)	1.3 (0.75)*
3. Active symptoms of major mental illness	0.26 (0.59)	0.19 (0.45)
4. Impulsivity	1.7 (0.51)	1.3 (0.74)**
5. Unresponsive to treatment	1.1 (0.70)	1.1 (0.59)
Risk management items		
1. Plans lack feasibility	1.1 (0.68)	1.2 (0.74)
2. Exposure to destabilizers	1.4 (0.54)	1.4 (0.62)
3. Lack of personal support	1.3 (0.66)	1.2 (0.70)
4. Noncompliance with remediation attempts	1.0 (0.63)	1.1 (0.65)
5. Stress	1.8 (0.38)	1.9 (0.26)
Historical scale	14.0 (2.9)	14.9 (3.0)
Clinical scale	5.4 (2.0)	5.4 (2.3)
Risk management scale	6.6 (1.9)	6.8 (2.1)
Total score	25.9 (5.5)	27.1 (6.5)
PCL-R		
Factor 1	6.1 (2.9)	7.8 (3.9)*
Factor 2	8.5 (4.2)	10.0 (4.6)
Total score	16.5 (6.2)	19.4 (8.5)
Final risk judgments and diagnosis of psychopathy		
	N (%)	N (%)
HCR-20: Low	11 (26%)	11 (26%)
HCR-20: Moderate	21 (50%)	13 (31%)*
HCR-20: High	10 (24%)	18 (43%)*
PCL-R \geq 26	4 (10%)	10 (24%)

* $p < 0.01$. ** $p < 0.05$ (two tailed).

specific risk factors that do not fit within the item descriptions. The HCR-20 has to be coded by an experienced forensic clinician, who should use all available information on the offender, preferably from different sources and gathered with different methods, for example, criminal records/police files, psychological reports, interviews with significant others and behavioral observations. The final risk judgment has to be indicated as low, moderate or high and is valid for a specific time period, for instance, within a specific setting or for a set time frame. The final risk judgment depends not only on the mere summation of the item scores, but also

on specific combinations of factors or other considerations. In some cases, only one or two items may be sufficient to justify the judgment 'high risk', for example, when a patient has florid psychotic symptoms (e.g. auditory command hallucinations that instruct the patient to commit homicide). The final risk judgment can be considered as a structured professional judgment that is arrived at through the process of coding the checklist and integrating all available information.

In the present study, the Dutch authorized adaptation of the HCR-20 was used. Prospective research with this Dutch version in a sample of

60 patients has demonstrated good inter-rater reliability (de Vogel & de Ruiter, 2004). In a retrospective study in a sample of 120 forensic psychiatric patients, de Vogel and colleagues (2004) found good predictive validity for the HCR-20 subscales, the total score and the final risk judgment.

Psychopathy Checklist-Revised

The PCL-R was designed to assess the construct of psychopathy and comprises two factors: Factor 1, which has been labeled selfish, callous and remorseless use of others, and Factor 2, which represents a chronically unstable and antisocial lifestyle (Hare, 1991). More recently, Cooke and Michie (2001) have subjected the PCL-R items to Item Response Theory (IRT) analyses and demonstrated that a hierarchical three-factor model (interpersonal, affective and behavioral factors) provides an even better understanding of the multifaceted concept of psychopathy. The instrument consists of 20 items that have to be coded on a three-point scale, '0', item does not apply, '1', the item probably or partially applies, and '2', the item definitely applies, on the basis of a semi-structured interview and collateral information. The total score can range from 0 to 40 and reflects an estimate of the degree to which an individual matches the prototypical psychopath. The cut-off score for the diagnosis of psychopathy is generally set at 30, but in several European countries, for instance Scotland, England, Sweden, and The Netherlands, a cut-off score of 25 or 26 has proven useful (Hare, Clark, Grann, & Thornton, 2000; Hildebrand, de Ruiter, & de Vogel, 2004). Although originally not developed as a risk assessment instrument, two meta-analyses showed the PCL-R to be strongly linked to repeated violent behavior in (mainly) male samples (Hemphill, Templeman, Wong, & Hare, 1998; Salekin, Rogers, & Sewell, 1996) and, therefore, psychopathy as measured by the PCL-R is included as one of the risk factors in risk assessment instruments such as the HCR-20. The ability of the PCL-R to predict recidivism was shown to possess cross-cultural generalizability (Hare et al., 2000). In the present study, the Dutch version of the PCL-R was used. Research in the Dr. Henri van der Hoeven Kliniek rendered good inter-rater reliability for this Dutch version (Hildebrand, de Ruiter, de Vogel, & van der Wolf, 2002). Furthermore, PCL-R scores were significantly related to disruptive behavior in a sample of 92 male forensic psychiatric inpatients (Hildebrand, de Ruiter, & Nijman, 2004).

Violent Outcome Data

Violent outcome data were obtained from two sources. First, data on violent recidivism of the patients from the retrospective study were retrieved from the Judicial Documentation register of the Ministry of Justice. For the identification of violent offenses, we adopted the HCR-20 definition of violence: 'violence is actual, attempted, or threatened harm to a person or persons' (Webster et al., 1997, p. 24). Second, data on inpatient violence were obtained from information bulletins that are published daily in the hospital to inform patients and staff. In these bulletins, the most important events of the day are reported, such as disruptive incidents that occurred during the last 24 hours, or positive results on urine analysis to detect whether a patient has taken drugs. Disruptive incidents are registered and assigned to one of four categories: verbal violence, verbal threat, physical violence, and violation of hospital rules (see for details Hildebrand, de Ruiter, & Nijman, 2004). Because the HCR-20 is designed to assess risk for violence to others, we only used the category physical violence, and only those incidents of physical violence directed towards other persons (e.g. staff or patients). For instance, property damage was not included, unless the property damage occurred in the presence of someone with the goal to frighten or threaten that person (e.g. smashing a cup of hot coffee against the wall while someone is standing close by). HCR-20 scores and final risk judgments were related to incidents of physical violence during treatment that occurred after the date of the risk assessment. Inpatient violence and violent recidivism after discharge were collapsed into one violent outcome variable.

Statistical Analyses

The *F*-test was used to examine differences between men and women in HCR-20 and PCL-R mean scores. For differences in HCR-20 final risk judgments and psychopathy diagnoses (PCL-R \geq 26) we used chi-square analysis. The interrater reliability of the HCR-20 was examined by means of the intraclass correlation coefficient (ICC), using the two-way random effect variance model and consistency type (McGraw & Wong, 1996). Critical values used for single measure ICCs are ICC \geq 0.75, excellent; $0.60 \leq$ ICC $<$ 0.75, good; $0.40 \leq$ ICC $<$ 0.60, moderate; ICC $<$ 0.40, poor (Fleiss, 1986). The predictive validity was established with receiver operating characteristic (ROC) analyses (see Douglas & Weir, 2003). The major advantage of this statistical method is its insensitivity to base rates. The ROC

analyses result in a plot of the true positive rate (sensitivity) against the false positive rate (one minus specificity) for every possible cut-off score of the instrument. The area under the curve (AUC) can be interpreted as the probability that a randomly selected recidivist would score higher on the instrument than a randomly selected non-recidivist. An AUC of 0.00 represents perfect negative prediction, an AUC of 0.50 chance prediction, and an AUC of 1.0 perfect positive prediction. In general, AUC values of 0.70 and above are considered as moderate, and values above 0.75 as good (Douglas & Weir, 2003). To compare the obtained AUC values for men and women, we used ACCUROC Version 2.5 (Vida, 1997) that applies the non-parametric method as described by DeLong, DeLong, and Clarke-Pearson (1988). Pearson point-biserial correlations were computed for comparative purposes.

RESULTS

Inter-rater Reliability

The inter-rater reliability for female patients was good for the Historical scale, total score, and final risk judgment ($N = 27$; ICC = 0.82, 0.75, and 0.74, respectively), and moderate for the Clinical scale and Risk management scale (ICC = 0.55, and 0.51, respectively). Furthermore, for male patients we found good inter-rater reliability for the Historical scale, Clinical scale, total score, and final risk judgment ($N = 28$; ICC = 0.82, 0.70, 0.77, and 0.69, respectively), and moderate inter-rater reliability for the Risk management scale (ICC = 0.49).

Risk Judgments and Psychopathy

Table 2 presents the mean scores and standard deviations for the HCR-20 items, subscales and total score for both female and male patients. As can be seen from this table, the mean HCR-20 subscales and total scores did not differ significantly between the female and male samples. However, there were significant differences on some individual HCR-20 item scores. Female patients received significantly lower scores on the items 'Young age at first violent incident', 'Psychopathy' and 'Negative attitudes'. In contrast, female patients compared with male patients received significantly higher scores on the items 'Relationship instability' and 'Impulsivity'. Regarding the HCR-20 final risk judgments, women were significantly more often

judged as moderate risk, while men were significantly more often judged as high risk. The mean HCR-20 total score per final risk judgment category for female patients was low, 21.6 (range = 10–29); moderate, 26.2 (range = 19–32); high, 30.2 (range = 23–37). For male patients, the mean HCR-20 total score per final risk judgment category was low, 20.5 (range = 12–29); moderate, 26.1 (range = 19–33); high, 31.8 (range = 21–37). For both men and women the mean HCR-20 total scores differed significantly between the low, moderate and high risk cases (women, $F = 7.8$, $p < 0.01$; men, $F = 17.2$, $p < 0.001$). There were no significant differences between men and women in the mean HCR-20 total scores per final risk judgment (low, $F = 0.04$, $p = 0.67$; moderate, $F = 0.29$, $p = 0.94$; high, $F = 0.31$, $p = 0.33$). Frequently coded 'other considerations' differed somewhat for female and male patients. The three most frequently coded 'other considerations' for male patients were financial problems (6), lack of prospects for the future (5) and violent fantasies (4) (number of codings for females 2, 3 and 2, respectively). The three most frequently coded 'other considerations' for female patients were forming a new intimate relationship (e.g. problematic partner choice) (18), care for children (5) and prostitution (4) (number of codings for males 2, 1 and 1, respectively).

The mean PCL-R Factor 1, Factor 2, and total score and the categorical diagnosis of psychopathy (PCL-R ≥ 26) are shown in Table 2. Female patients compared with males received significantly lower mean scores on Factor 1 ($F = 3.7$, $p < 0.05$); however, the differences in the mean Factor 2 and total score were not or only marginally significant ($F = 1.1$, $p = 0.11$; $F = 5.1$, $p = 0.08$, respectively). Male patients compared with females were more often diagnosed as psychopathic (PCL-R ≥ 26), although this difference was marginally significant ($\chi^2 = 3.1$, $p = 0.08$).

Violent Outcome

First, violent reconvictions after discharge from the hospital were calculated for the 15 female and 21 male patients from the retrospective study. Significantly more male ex-patients compared with female ex-patients were convicted for a violent reoffense: nine (43%) of 21 males versus two (13%) of 15 females ($\chi^2 = 3.6$, $p < 0.05$; odds ratio = 4.9, 95% CI = 1.1–32.9). Next, we computed the number of patients who had been physically violent towards others during their stay in the hospital for the 27 female and 21 male patients from the

prospective study. There was no significant difference between women and men: eight (30%) of 27 female patients were registered to have been physically violent during their hospital stay, versus six (29%) of 21 male patients. Examples of physical violence during treatment were throwing hot coffee at staff, hitting staff or fellow patients, and seizing someone by the throat.

Predictive Validity

Table 3 shows the AUC values and Pearson correlations of the HCR-20 subscales and total scores for both female and male patients regarding violent outcome, and Figures 1 and 2 present the ROC curves for the HCR-20 for violent outcome. For female patients, only the AUC value of the HCR-20 final risk judgment was significantly above 0.50. Similarly, only the correlation between the HCR-20 final risk judgment and violent outcome was significant. The difference in violent outcomes between female patients who were judged to pose a low, moderate or high risk was significant ($\chi^2 = 16.2$, $p < 0.001$, violent outcome 0, 14 and 77%, respectively). Female patients who scored above the median (HCR-20 total score = 26.6) compared with those who scored below did not show significantly more violent outcomes (29 versus 19%).

For male patients, the AUC values for violent outcome were significantly above 0.50 for all HCR-20 subscales, the total score and final risk

judgment. Also, Pearson correlations between the HCR-20 subscale scores, total score, final risk judgment and violent outcome were significantly positive. The difference in violent outcomes between male patients who were judged to pose a low, moderate or high risk was significant ($\chi^2 = 24.4$, $p < 0.001$, violent outcome 0, 8 and 78%, respectively). Male patients who scored above the median (HCR-20 total score = 28.5) compared with those who scored below showed significantly more violent outcomes ($\chi^2 = 12.5$, $p < 0.001$, 62 versus 10%). When we compared the AUC values for violent recidivism after discharge with AUC values for inpatient violence, we found no substantial differences in predictive accuracy for either the female or male sample. The HCR-20 Risk management scale and total score were significantly more accurate in predicting violent recidivism in men than in women (Z -statistic = 2.9 and 2.5, respectively, $p < 0.01$, two tailed).

Table 3 also presents the AUC values and Pearson correlations of the PCL-R factors and total scores for both female and male patients regarding violent outcome. For the female patients, none of the AUC values or Pearson correlations were significant. For the male patients, AUC values for violent outcome were significantly above 0.50 for the PCL-R Factor 2 and total score, and correlations between violent outcome and PCL-R Factor 2 score and total score were significant. The PCL-R Factor 1, Factor 2 and total score were significantly more

Table 3. Predictive validity of the HCR-20 and PCL-R for female and male patients

	Violent outcome for females <i>N</i> = 42			Violent outcome for males <i>N</i> = 42		
	AUC	SE	<i>r</i>	AUC	SE	<i>r</i>
HCR-20						
Historical scale	0.63	0.11	0.22	0.83**	0.06	0.54*
Clinical scale	0.61	0.10	0.17	0.75**	0.08	0.42*
Risk management scale	0.52	0.11	0.07	0.88**	0.05	0.62*
Total score	0.59	0.11	0.20	0.88**	0.05	0.59*
Final risk judgment	0.86**	0.07	0.57*	0.91**	0.05	0.70*
PCL-R						
Factor 1	0.36	0.08	-0.21	0.64	0.09	0.24
Factor 2	0.41	0.11	-0.10	0.84**	0.06	0.58*
Total score	0.34	0.10	-0.21	0.74*	0.08	0.42*
PCL-R \geq 26	0.50	0.11	0.01	0.63	0.09	0.28

* $p < 0.01$, ** $p < 0.001$ (two tailed).

AUC = area under the curve. SE = standard error. r = Pearson point-biserial correlation coefficient. Violent outcome = inpatient violence or violent recidivism after discharge.

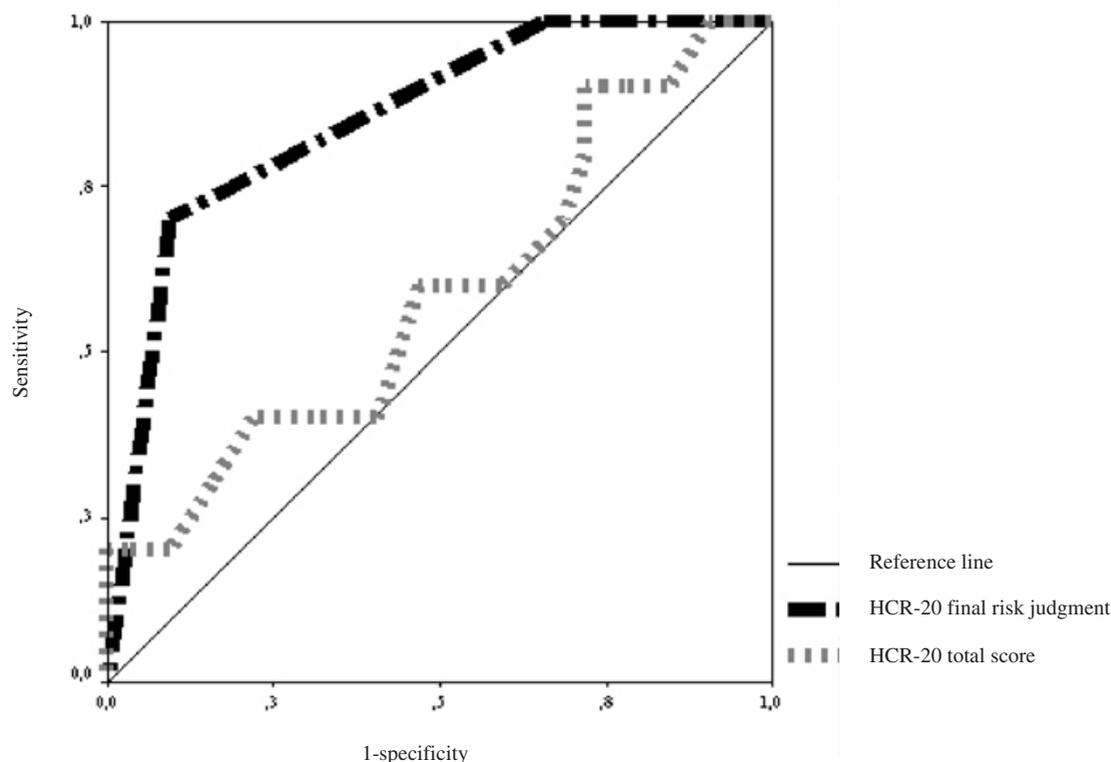


Figure 1. ROC curves of HCR-20 total score and final risk judgment for violent outcome in a sample of 42 female patients

accurate in predicting violent recidivism in men than in women (Z -statistic = 2.3, 3.4 and 3.1, respectively, $p < 0.01$, two tailed).

DISCUSSION

In this study, a sample of 42 female forensic psychiatric patients was compared with a matched sample of 42 male forensic psychiatric patients on base rates of violent outcome, HCR-20 and PCL-R scores and predictive validity of the latter instruments. We found several significant differences between women and men, most importantly in the predictive validity of the HCR-20 and PCL-R, but also in mean HCR-20 individual item scores, base rates for violence after discharge from the hospital and sample characteristics.

First, we found some significant differences in sample characteristics, despite our matching procedure. Female patients more often had a diagno-

sis of borderline personality disorder and less often a narcissistic personality disorder or antisocial personality disorder. This is in line with the study by Strand and Belfrage (2001) and also with research that suggests that borderline personality disorder is much more common among women (Weisman, 1993). We believe the large proportion of female borderline patients has a considerable impact on the interpretation of our results, for instance, on differences in HCR-20 item scores (see below). Furthermore, women obtained higher scores on intelligence scales, especially on verbal intelligence, were significantly older at the time of their first conviction, and—albeit not significantly—had fewer previous convictions than men.

Second, there were no significant differences in mean HCR-20 subscale and total scores for male and female patients, and this finding resembles those of Strand and Belfrage (2001) and Nicholls (unpublished master's thesis). Our finding that female patients had significantly lower mean scores

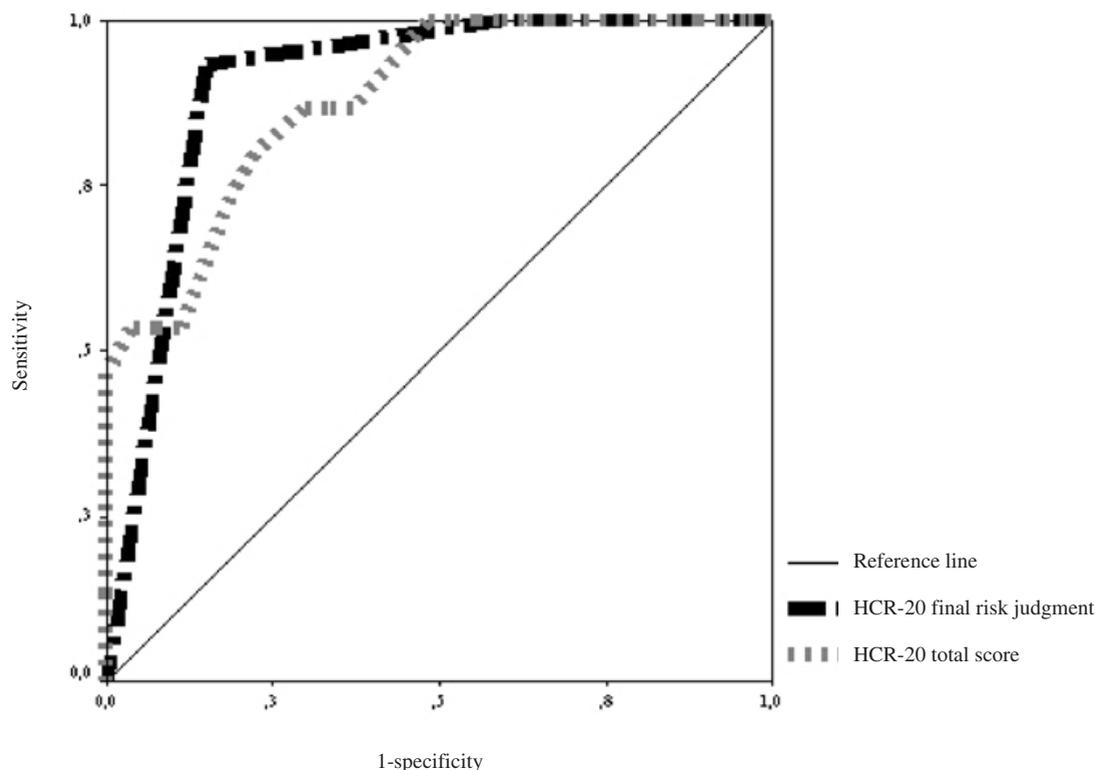


Figure 2. ROC curves of HCR-20 total score and final risk judgment for violent outcome in a sample of 42 male patients

on 'Young age at first violent incident' and 'Negative attitudes' and significantly higher mean scores on 'Impulsivity' is in line with the work of Strand and Belfrage (2001). In the present study, we also found significantly lower mean scores on 'Psychopathy' and significantly higher mean scores on 'Relationship instability' for women. An explanation for the higher mean scores on 'Relationship instability' and 'Impulsivity' could be that both factors are criteria for the borderline personality disorder diagnosis, which was highly prevalent in our female sample. The lower mean score on 'Young age at first violent incident' is in accordance with previous research that showed a later onset of criminal behavior in girls as compared with boys (Silverthorn & Frick, 1999). The lower score on 'Negative attitudes' could be explained by the fact that in general women have different motives for their violent offenses compared to men, more often reactive and relational, and less instrumental or resulting from criminogenic needs (see Crick &

Grottpeter, 1995). This hypothesis is confirmed by the lower prevalence of antisocial personality disorder we found in our female patients. The lower mean score on the item 'Psychopathy' is in line with the lower mean scores on the PCL-R, although the differences in mean Factor 2 scores and total scores were not or only marginally significant. The lower prevalence of psychopathy among female patients is in line with previous research into psychopathy in females (Grann, 2000; Salekin et al., 1997; Vitale et al., 2002; Warren et al., 2003).

Third, regarding violent outcome we found the base rate for inpatient violence to be similar for female and male patients. This was also demonstrated in other studies (Lidz et al., 1993; Nicholls, unpublished matter's thesis). The base rates for inpatient violence in our study (women 30%; men 29%) are similar to the base rates for physical inpatient aggression (women 30%; men 27.5%) found in Nicholls' study (unpublished matter's thesis). Male patients were found to be five times more

likely to be convicted for a violent reoffense after discharge from the hospital than female patients. This finding resembles the finding of Ross et al. (1998), who found male patients to be four times more likely than female patients to express any aggression. Thus, although there was no gender difference in the base rate of inpatient violence, base rates for community violence were significantly different. A possible explanation for this difference is that female violence in the community is often less visible and more subtle or manipulative, for instance, in domestic violence or child abuse. Research has demonstrated that the prevalence rate of domestic violence by women is comparable to or even higher than the prevalence rate of domestic violence by men (Magdol et al., 1997). Domestic violence is less likely to come to the attention of the criminal justice system than violence committed in the public environment, which is much more commonly committed by men. Moreover, the police often respond differently to violence when it is committed by a female perpetrator versus a male perpetrator. Pajer (1998) has described this gender bias in the justice system, i.e. the reluctance to arrest women coupled with a tendency toward psychiatric referrals.

Fourth, the interrater reliability of the HCR-20 in the present study was in line with previous studies (see Douglas & Weir, 2003). We found no substantial differences in inter-rater reliability between men and women.

Finally, we found poor predictive validity for the HCR-20 numerical scores for women compared with good to excellent predictive validity for men. Notable, however, was the good predictive validity of the HCR-20 final risk judgment for both female (AUC = 0.86) and male patients (AUC = 0.91). Thus, while a simple addition of individual HCR-20 risk factors was not adequate in predicting violence risk in our female patients, the SPJ method based on the HCR-20 seemed to perform well. For our male sample the structured final risk judgment yielded the highest AUC value, and this is in line with previous research that demonstrated the structured final risk judgment to add incremental validity to the HCR-20 total score used in an actuarial sense (Douglas, Ogloff, & Hart, 2003). The same was found for the Spousal Assault Risk Assessment guide (SARA; Kropp, Hart, Webster, & Eaves, 1999), an SPJ guideline for the assessment of domestic violence (Kropp & Hart, 2000). The poor predictive validity found for the HCR-20 total score in female patients is in contrast with the results of Nicholls (unpublished master's thesis).

A possible explanation is the difference in the samples that were studied. The patients in Nicholls' sample were mainly suffering from Axis I disorders (87%), and only four percent received a diagnosis of borderline personality disorder. This is in sharp contrast to our sample, where three-quarters of the women suffered from borderline personality disorder and Axis I disorders were usually not the primary diagnosis. Furthermore, Nicholls used a different definition of violence; for instance, she also considered property damage and verbal aggression, while in our study we limited ourselves to physical violence towards others. In our study, the PCL-R was demonstrated to be a good predictor of violence for male patients, but not for female patients. This finding is in line with previous studies that found good predictive validity for future violence in (mainly) male samples (Hemphill et al., 1998; Salekin et al., 1996), but modest predictive validity for future violence in female samples (see Vitale & Newman, 2001). Thus, the results of our study suggest that the PCL-R is not a valid assessment of the psychopathy construct in Dutch female forensic psychiatric patients.

A number of limitations to the present study should be mentioned. First of all, the design of the study was mixed, because we combined patients and violent outcome data from a retrospective study and a prospective study. The reason for mixing the patients from a retrospective and a prospective design was to obtain a large enough sample. In The Netherlands, women make up only five percent of the tbs-population. Second, the violent outcome data may have been an underestimate of actual violence. The violent recidivism data were retrieved from only one source, the Judicial Documentation register of the Ministry of Justice. As a consequence, the reconviction rate is inevitably an underestimation of the actual recidivism rate, because not all offenders are reported, apprehended and arrested. With regard to the prospective outcome data, incidents of physical violence are not always reported on the information bulletins. For example, it is possible that incidents of physical violence between patients are not observed by staff or reported by patients to staff. Third, the sample sizes were relatively small and only derived from one site. Larger samples would have resulted in increased power. However, given that there is such a paucity of research on female forensic psychiatric patients, we believe that even matched samples of limited size such as ours can make a contribution to the knowledge base.

Our findings demonstrate that the method of structured professional judgment, i.e. systematically rating risk factors, integrating and weighing information, is effective in both male and female patients. For research purposes, we recommend researchers who conduct studies in mixed gender samples to report the results on predictive validity of risk assessment instruments separately for men and women, because reporting the results jointly could lead to distorted conclusions. Perhaps in different patients the HCR-20 will show good predictive validity, as in Nicholls' study with primarily Axis I disordered women or in civil psychiatric samples (see Nicholls, Ogloff, Douglas, & Grant, 2004). Risk assessment research in female forensic psychiatric patients is still a relatively unexplored area. Although women are only a minority in forensic psychiatry, it seems that in the past two decades female aggression is on the rise, especially among young girls (English, 1993; Mertens, Grapendaal, & Docter-Schamhardt, 1998; Odgers & Moretti, 2002). More knowledge on specific violence risk factors in women and the risk management strategies needed to prevent repeated violence in women is desirable. This is also important from a public mental health perspective because research has demonstrated an intergenerational transfer of risk of aggression between mothers and children; mothers with a history of violent offense(s) more often have disruptive, aggressive children (Serbin et al., 1998).

REFERENCES

- American Psychiatric Association (APA). (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Archer, D., & McDaniel, P. (1995). Violence and gender. In R.B. Ruback, & N.A. Weiner (Eds), *Interpersonal violent behaviors: Social and cultural aspects* (pp. 63–87). New York: Springer.
- Blanchette, K. (1994). Classifying female offenders for correctional interventions. *Forum on Corrections Research*, 6, 36–41.
- Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines and training. *American Psychologist*, 51, 945–956.
- Chesney-Lind, M. (1989). Girls' crime and woman's place: Toward a feminist model of female delinquency. *Crime and Delinquency*, 35, 5–29.
- Coid, J., Kahtan, N., Gault, S., & Jarman, B. (1999). Patients with personality disorder admitted to secure forensic psychiatric services. *British Journal of Psychiatry*, 175, 528–536.
- Cooke, D., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment*, 13, 171–188.
- Crick, N.R., & Grotpeter, J.K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development*, 66, 710–722.
- de Ruiter, C., & Greeven, P.G.J. (2000). Personality disorders in a Dutch forensic psychiatric sample: Convergence of interview and self-report measures. *Journal of Personality Disorders*, 14, 162–170.
- de Vogel, V., & de Ruiter, C. (2004). Differences between clinicians and researchers in assessing risk of violence in forensic psychiatric patients. *The Journal of Forensic Psychiatry and Psychology*, 15, 145–164.
- de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B., & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20. A retrospective study in a Dutch sample of forensic psychiatric patients. *International Journal of Forensic Mental Health*, 3, 149–165.
- DeLong, E.R., DeLong, D.M., & Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristics curves: A non-parametric approach. *Biometrics*, 44, 837–854.
- Douglas, K.S., Ogloff, J.R.P., & Hart, S.D. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services*, 54, 1372–1379.
- Douglas, K.S., & Weir, J. (2003). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*. Retrieved October 8, 2004, from <http://www.sfu.ca/psychology/groups/faculty/hart/violink.htm>
- English, K. (1993). Self-reported crime rates of women prisoners. *Journal of Quantitative Criminology*, 9, 357–381.
- Fleiss, J.L. (1986). *The design and analysis of clinical experiments*. New York: Wiley.
- Funk, S.J. (1999). Risk assessment for juveniles on probation. A focus on gender. *Criminal Justice and Behavior*, 26, 44–68.
- Grann, M. (2000). The PCL-R and gender. *European Journal of Psychological Assessment*, 16, 147–149.
- Hare, R.D. (1991). *Manual for the Hare Psychopathy Checklist-Revised*. Toronto: Multi-Health Systems.
- Hare, R.D., Clark, D., Grann, M., & Thornton, D. (2000). Psychopathy and the predictive validity of the PCL-R: An international perspective. *Behavioral Sciences and the Law*, 18, 623–645.
- Harer, M.D., & Langan, N.P. (2001). Gender differences in predictors of prison violence: assessing the predictive validity of a classification system. *Crime and Delinquency*, 47, 513–536.
- Hart, S.D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology*, 3, 121–137.
- Hart, S.D., Cox, D., & Hare, R.D. (1995). *The Hare Psychopathy Checklist: Screening Version (PCL:SV)*. Toronto: Multi-Health Systems.
- Hemphill, J.F., Templeman, R., Wong, S., & Hare, R.D. (1998). Psychopathy and crime: Recidivism and criminal careers. In D.J. Cooke, A.E. Forth, & R.D. Hare (Eds), *Psychopathy: Theory, research and implications for society* (pp. 375–399). Dordrecht: Kluwer.

- Hildebrand, M., & de Ruiter, C. (2004). PCL-R psychopathy and its relation to DSM-IV Axis I and Axis II disorders in a sample of male forensic psychiatric patients in the Netherlands. *International Journal of Law and Psychiatry*, 27, 233–248.
- Hildebrand, M., de Ruiter, C., & de Vogel, V. (2004). Psychopathy and sexual deviance in treated rapists: Association with sexual and nonsexual recidivism. *Sexual Abuse: A Journal of Research and Treatment*, 16, 1–24.
- 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.
- Hildebrand, M., de Ruiter, C., & Nijman, H. (2004). PCL-R psychopathy predicts disruptive behavior among male offenders in a Dutch forensic psychiatric hospital. *Journal of Interpersonal Violence*, 19, 13–29.
- Kropp, P.R., & Hart, S.D. (2000). The Spousal Assault Risk Assessment (SARA) guide: Reliability and validity in adult male offenders. *Law and Human Behavior*, 24, 101–118.
- Kropp, P.R., Hart, S.D., Webster, C.D., & Eaves, D. (1999). *Manual for the Spousal Assault Risk Assessment Guide (Version 3)*. Vancouver: British Columbia Institute against Family Violence.
- Laws, D.R., Hudson, S.M., & Ward, T. (2000). *Remaking relapse prevention: A sourcebook*. London: Sage.
- 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: Earls court Child and Family Centre.
- Lidz, C.W., Mulvey, E.P., & Gardner, W. (1993). The accuracy of predictions of violence to others. *Journal of the American Medical Association*, 269, 1007–1011.
- Loucks, A.D., & Zamble, E. (1999). Canada searches for predictors common to both men and women. *Corrections Today*, 61, 26–32.
- Magdol, L., Moffitt, T.E., Caspi, A., Newman, D.L., Fagan, J., & Silva, P.A. (1997). Gender differences in partner violence in a birth cohort of 21 years olds: Bridging the gap between clinical and epidemiological approaches. *Journal of Consulting and Clinical Psychology*, 65, 68–78.
- McGraw, K.O., & Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30–46.
- McNiel, D.E., & Binder, R.L. (1990). The relationship of gender to violent behavior in acutely disturbed psychiatric patients. *Journal of Clinical Psychiatry*, 51, 110–114.
- McNiel, D.E., & Binder, R.L. (1995). Correlates of accuracy in the assessment of psychiatric inpatients' risk of violence. *American Journal of Psychiatry*, 152, 901–906.
- Mertens, N.M., Grapendaal, M., & Docter-Schamhardt, B.J.W. (1998). *Delinquency in girls in The Netherlands (Report No. 169)*. The Hague: Ministry of Justice, Scientific Research and Documentation Centre.
- Monahan, J. (1981). *The clinical prediction of violent behavior*. Rockville, MD: National Institute of Mental Health.
- Monahan, J., Steadman, H.J., Silver, E., Appelbaum, P.S., Robbins, P.C., Mulvey, E.P., Roth, L.H., Grisso, T., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. Oxford: Oxford University Press.
- Newhill, C.E., Mulvey, E.P., & Lidz, C.W. (1995). Characteristics of violence in the community by female patients seen in a psychiatric emergency service. *Psychiatric Services*, 46, 785–789.
- Nicholls, T.L., Ogloff, J.R.P., Douglas, K.S., & Grant, I. (2004). Assessing risk for violence among male and female civil psychiatric patients: The HCR-20, PCL: SV, and McNiel & Binder's Screening Measure. *Behavioral Sciences and the Law*, 22, 127–158.
- Ogders, C.L., & Moretti, M.M. (2002). Aggressive and antisocial girls: Research update and challenges. *International Journal of Forensic Mental Health*, 1, 103–119.
- Pajer, K.A. (1998). What happens to 'bad' girls? A review of the adult outcomes of antisocial adolescent girls. *American Journal of Psychiatry*, 155, 862–870.
- Pfohl, B., Blum, N., & Zimmerman, M. (1995). *Structured Interview for DSM-IV Personality, SIDP-IV*. Iowa: University of Iowa.
- Philipse, M., de Ruiter, C., Hildebrand, M., & Bouman, Y. (2000). *HCR-20. Beoordelen van het risico van geweldadig gedrag. Versie 2 [HCR-20. Assessment of the risk of violent behavior. Version 2]*. Nijmegen: Pompestichting—van der Hoeven Stichting.
- Quinsey, V.L., Harris, G.T., Rice, M.E., & Cormier, C.A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Ross, D.J., Hart, S.D., & Webster, C.D. (1998). *Aggression in psychiatric patients. Using the HCR-20 to assess risk for violence in hospital and in the community*. Port Coquitlam, BC: Riverview Hospital.
- Salekin, R.T., Rogers, R., & Sewell, K.W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist-Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice*, 3, 203–215.
- Salekin, R.T., Rogers, R., & Sewell, K.W. (1997). Construct validity of psychopathy in a female offender sample: A multitrait—multimethod evaluation. *Journal of Abnormal Psychology*, 106, 576–585.
- Scarth, K., & McLean, H. (1994). The psychological assessment of women in prison. *Forum on Corrections Research*, 6, 32–35.
- Serbin, L.A., Cooperman, J.M., Peters, P.L., Lehoux, P.M., Stack, D.M., & Schwartzman, A.E. (1998). Intergenerational transfer of psychosocial risk in women with childhood histories of aggression, withdrawal, or aggression and withdrawal. *Developmental Psychology*, 34, 1246–1262.
- Silverthorn, P., & Frick, P.J. (1999). Developmental pathways to antisocial behavior: The delayed-onset pathway in girls. *Development and Psychopathology*, 11, 101–126.
- Simourd, L., & Andrews, D.A. (1994). Correlates of delinquency: A look at gender differences. *Forum on Corrections Research*, 6, 26–31.
- Strand, S., & Belfrage, H. (2001). Comparison of HCR-20 scores in violent mentally disordered men and women: Gender differences and similarities. *Psychology, Crime, and Law*, 7, 71–79.

- Tardiff, K., Marzuk, P.M., Leon, A.C., Portera, L., & Weiner, C. (1997). Violence by patients admitted to a private psychiatric hospital. *American Journal of Psychiatry*, *154*, 88–93.
- Vertommen, H., Verheul, R., de Ruiter, C., & Hildebrand, M. (2002). *De herziene versie van Hare's Psychopathie Checklist [Revised version of Hare's Psychopathy Checklist]*. Lisse: Swets.
- Vida, S. (1997). *AccuROC nonparametric receiver operating characteristic analysis (Version 2.5)* [Computer Software]. Montreal, Quebec, Canada: Accumetric Corporation.
- Vitale, J.E., & Newman, J.P. (2001). Using the Psychopathy Checklist–Revised with female samples: Reliability, validity and implications for clinical utility. *Clinical Psychology: Science and Practice*, *8*, 117–132.
- Vitale, J.E., Smith, S.S., Brinkley, C.A., & Newman, J.P. (2002). The reliability and validity of the Psychopathy Checklist–Revised in a sample of female offenders. *Criminal Justice and Behavior*, *29*, 202–231.
- Warren, J.I., Burnette, M.L., South, S.C., Chauhan, P., Bale, R., & Friend, R. (2002). Personality disorders and violence among female prison inmates. *The Journal of the American Academy of Psychiatry and the Law*, *30*, 502–509.
- Warren, J.I., Burnette, M.L., South, S.C., Chauhan, P., Bale, R., Friend, R., & van Patten, I. (2003). Psychopathy in women: Structural modeling and comorbidity. *International Journal of Law and Psychiatry*, *26*, 223–242.
- Webster, C.D., Douglas, K.S., Eaves, D., & Hart, S.D. (1997). *HCR-20. Assessing the risk of violence. Version 2*. Burnaby, British Columbia: Simon Fraser University and Forensic Psychiatric Services Commission of British Columbia.
- Weisman, M.M. (1993). The epidemiology of personality disorders: A 1990 update. *Journal of Personality Disorders*, *7S*, 44–62.