Psychology, Crime & Law, June 2006; 12(3): 321–336

Routledge
Taylor & Francis Group

ORIGINAL ARTICLE

Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20

VIVIENNE DE VOGEL[1] & CORINE DE RUITER[2,3]

[1]*Dr Henri van der Hoeven Kliniek, Department of Research, Dr Henri van der Hoeven Kliniek, PO Box 174, 3500 AD Utrecht, The Netherlands,* [2]*Department of Psychology, Maastricht University, and* [3]*Netherlands Institute of Mental Health and Addiction (Trimbos Institute), PO Box 725, 3500 AS Utrecht, The Netherlands*

(Received 29 March 2005; accepted 10 January 2006)

Abstract
In this prospective study, the Dutch version of the HCR-20 (an instrument assessing risk factors for violence in the past, present and future) was coded independently by three rater groups (researchers, treatment supervisors and group leaders) for 127 male mentally disordered offenders admitted to a forensic psychiatric hospital. During case conferences, the three raters discussed their ratings and reached consensus on their ratings and final risk judgment. HCR-20 ratings were related to incidents of physical violence during treatment. Overall, the predictive validity of the HCR-20 was good. We found no differences between researchers and treatment supervisors in predictive accuracy. Group leaders performed worse compared to the other two rater groups. The consensus rating was the best predictor. Implications for structured violence risk assessment in clinical practice are discussed.

Keywords: *Risk assessment, violence, HCR-20, predictive validity*

Introduction

During the last two decades, research into risk factors for violence, the development of risk assessment instruments and research into the psychometric properties of these instruments has expanded enormously. To date, numerous structured risk assessment instruments are available for mental health professionals working in forensic or general psychiatry or in the penitentiary system. Risk assessment instruments can be divided into strictly actuarial and structured professional judgment (SPJ) instruments. Actuarial instruments are developed solely based on risk factors that are empirically related to (sexually) violent behavior. These instruments are relatively simple to code – according to fixed rules and not necessarily by a

Correspondence: Vivienne de Vogel, PhD, Department of Research, Dr Henri van der Hoeven Kliniek, PO Box 174, 3500 AD Utrecht, The Netherlands. E-mail: vdevogel@hoevenstichting.nl

ISSN 1068-316X print/ISSN 1477-2744 online
DOI: 10.1080/10683160600569029

forensic expert – and contain predominantly static, historical factors that are added up according to an algorithm to reach a conclusion regarding the risk of recidivism. Examples of actuarial instruments are the Violence Risk Appraisal Guide (VRAG; Harris & Rice, 1997) for violent behavior, and the Static-2002 (Hanson & Thornton, 2002) for sexual violence. 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 (sexual) violence, static as well as dynamic factors. The essential difference between the actuarial and the SPJ approach is in how the final risk judgments are arrived at; in actuarial instruments by a fixed algorithm and in SPJ guidelines by (structured) human decision making.[1]

A SPJ risk assessment instrument that is internationally well known and the subject of numerous studies is the Historical, Clinical, Risk Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997). This instrument consists of 20 items representing risk factors for violence in the past, present and future. Research in various psychiatric and forensic settings in different countries has demonstrated good inter-rater reliability and predictive validity for the HCR-20 (Belfrage, 1998; Belfrage, Fransson, & Strand, 2000; Douglas, Ogloff, & Hart, 2003; Douglas, Ogloff, Nicholls, & Grant, 1999; Strand, Belfrage, Fransson, & Levander, 1999; see Douglas, Guy, & Weir, 2005 for a review of studies into the HCR-20). For instance, Douglas et al. (2003) found good predictive validity for the HCR-20 in a sample of 100 forensic psychiatric patients. Moreover, they demonstrated that the HCR-20 structured final risk judgment added incremental validity to the HCR-20 used in an actuarial sense, i.e. a simple addition of the scores on the 20 items.

An important limitation of many studies into the HCR-20 – which is designed for the prediction and management of *future* violence – is their retrospective design (see also Dernevik, 2004). Only a few prospective studies into the predictive validity of the HCR-20 have been published thus far (e.g. Belfrage et al., 2000; Dernevik, Grann, & Johansson, 2002; Dolan & Khawala, 2004). Another limitation of most studies of the HCR-20 concerns their ecological validity, i.e. their relevance to actual clinical risk assessment practice. In most published studies, the HCR-20 is coded by independent researchers, not by practicing clinicians. Generally, these researchers did not know the patients personally and coded the HCR-20 exclusively based on file information. Recently, Webster, Müller-Isberner, and Fransson (2002) referred to this problem and argued that "much more *in situ* research needs to be accomplished with instruments like the HCR-20" (p. 189).

When the HCR-20 is employed in clinical practice for the assessment of risk of future violence and in leave decision-making, ratings by experienced clinicians are required (Webster et al., 1997). Furthermore, in clinical practice it is customary that the treatment staff are responsible for leave decisions (Dernevik, Falkheim, Holmqvist, & Sandell, 2001). However, there is also some doubt about the objectivity of clinicians, especially clinicians who are closely involved in the treatment of the patient (for a more detailed discussion see: Dernevik et al., 2001; Litwack & Schlesinger, 1999; de Vogel & de Ruiter, 2004). Thus, the question arises of who is most likely to conduct accurate risk assessments: the objective, more distant researcher–assessor or the experienced clinician who knows the patient personally. Possibly, the consensus between the researcher and clinician will be the most accurate in predicting violence. To our knowledge, no studies have yet been published that examine differences in predictive accuracy of structured violence risk assessment instruments, such as the HCR-20, between clinicians and researchers or between individual ratings and consensus ratings by a group of raters. However, a few studies addressed the issue of multiple raters. McNiel, Lam, and Binder (2000) examined whether the predictive

accuracy of clinical assessments of violence risk improves when there is agreement between multiple clinicians (physicians and nurses). They found that when two clinicians reached similar conclusions these were more accurate than the conclusions of either clinician alone when their assessments disagreed. Huss and Zeiss (2004) found that individual clinicians demonstrated poor ability to predict violence among general psychiatric patients, but that the accuracy of the risk assessments improved much when they were aggregated as "group" decisions. It should be noted that in both of these studies, the clinicians did not actually meet and discuss; their ratings were aggregated by the research group. In conclusion, it is important to examine if there are differences between researchers and clinicians in the accuracy of their risk assessments and to compare this accuracy with the consensus between researchers and clinicians.

In this article, results of a prospective study are presented which started in January 2001. The authorized Dutch version of the HCR-20 (Philipse, de Ruiter, Hildebrand, & Bouman, 2000) was coded for 127 male patients admitted to the Dr Henri van der Hoeven Kliniek, a Dutch forensic psychiatric hospital, by both clinicians (group leaders and treatment supervisors) and independent researchers. In a previous study, we have examined the inter-rater reliability of the Dutch HCR-20 and differences between researchers and clinicians in coding the HCR-20 in 60 patients from this hospital, a subgroup of the present sample.[2] Overall, the inter-rater reliability of the HCR-20 was good. The group leaders gave significantly lower HCR-20 scores than the researchers. There were no significant differences between the mean HCR-20 scores of treatment supervisors and researchers, but there was a significant difference in the interpretation of the scores: treatment supervisors had more 'low risk' final judgments than researchers (de Vogel & de Ruiter, 2004). The goals of the present study were to establish the predictive validity of the Dutch HCR-20 and to gain insight into differences in risk assessment accuracy between (1) researchers, treatment supervisors and group leaders, and (2) individual ratings and consensus ratings. Also, following Douglas et al. (2003), we wanted to examine if the HCR-20 structured final risk judgment adds incremental validity to the HCR-20 actuarial score.

Method

Setting

This study was conducted at the Dr Henri van der Hoeven Kliniek, a forensic psychiatric hospital with 100 inpatient beds and 30 outpatient places in Utrecht, a city with 265 000 inhabitants in the center of The Netherlands. Patients are admitted under the judicial measure *terbeschikkingstelling* (tbs), which can be 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 because of severe psychopathology. The tbs order is of indefinite duration; every 1 or 2 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 inpatient forensic psychiatric institutions in The Netherlands. The hospital provides a varied treatment program, which includes job training, education, sports, creative arts, and psychotherapy. The treatment model of the hospital is cognitive–behavioral with an emphasis on relapse prevention. The "no cure but control" principle prevails (Laws, Hudson, & Ward, 2000). The emphasis of treatment is not on changing the personality of the offender, but on reducing/managing risk factors for recidivism. An important phase in the treatment is the transmural treatment phase. During this

resocialization phase, the patient lives outside the hospital, but is still undergoing treatment and is supervised by a specialized treatment team from the hospital.

Sample characteristics

The current sample included 127 men. The mean age at admission was 32.9 (SD = 9.6, range = 17–66). The majority of the patients were of Dutch nationality (80%). About half of the patients was unemployed (49%) and 60% were single at the time of the index offense. The majority of the patients had been convicted before their tbs order (77%) with an average number of 5.0 (SD = 6.1, range = 0–30) prior convictions. The index offenses were: 44% (attempted) homicide, 33% sexual offenses, 16% other violent offenses (e.g. robbery) and 7% arson. The mean length of stay in the hospital was 3.7 years (SD = 2.4, range = 0–12). More than half of the patients had abused substances in the past (8% alcohol, 15% drugs, and 44% multiple substances). In 5% of the patients, only an Axis I disorder (according to the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders*; DSM-IV; American Psychiatric Association, 1994) was diagnosed; 66% met the criteria for one or more Axis II disorders, particularly cluster B disorders[3] and in 28% there was comorbidity of Axis I and II disorders.[4] The majority of the patients had a history of psychiatric treatment; 49% had been admitted to a psychiatric institution and 24% had received outpatient treatment. The mean Psychopathy Checklist-Revised (PCL-R; Hare, 1991, 2003) total score of the patients was 21.5 (SD = 8.4, range = 2–38).

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, which relate to risk factors in the past, present and future, respectively (see Table III). The Historical items are static,[5] whilst the Clinical and Risk management factors are considered to be changeable, for instance, due to clinical intervention. The risk management items have to be coded with the context of the assessment – inside or outside an institution – in mind. 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 PCL-R or the Psychopathy Checklist: Screening Version (PCL:SV; Hart, Cox, & Hare, 1995) must have been administered to code the seventh item "Psychopathy". The structured final risk judgment has to be rated 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 is performed on a case by case basis and can be considered as a SPJ that is arrived at through the process of coding the checklist and integrating all available information. In arriving at the final risk judgment, the rater should also consider the degree of risk management that is necessary to prevent violence.

In the present study, the Dutch authorized adaptation of the HCR-20 was used. Earlier research in a sample of 60 patients from the Dr Henri van der Hoeven Kliniek demonstrated good inter-rater reliability (three raters, single measure Intraclass correlation coefficient (ICCs) varied from 0.57 to 0.82; de Vogel and de Ruiter, 2004). In a

retrospective study in a sample of 120 forensic psychiatric patients, good predictive validity was found for the HCR-20 subscales, total score and final risk judgment (de Vogel, de Ruiter, Hildebrand, Bos, & van de Ven, 2004).

Procedure

All raters were trained in coding the HCR-20 during a 1-day workshop given by a senior clinical psychologist and a research psychologist (the first author). In this workshop, the relevant empirical literature was discussed and the HCR-20 coding procedure was practiced using file information and videotapes of actual cases. Raters were instructed to use the HCR-20 manual and all available file information for all cases.

In the period of this study – 1 January 2001 until 1 June 2004 – the HCR-20 was coded for 127 patients who can be divided into different groups according to their treatment phase. In the course of treatment in the hospital, a number of specific phases can be distinguished in which the liberties of a patient can be increased and at that time the risk of violence needs to be (re-)evaluated. These phases are: when a patient has his first unsupervised leave from the hospital and when a patient is to enter the transmural treatment phase. The HCR-20 was coded for patients who were entering the above two phases ($n = 9$ and 28, respectively), and for patients who were already in the transmural treatment phase ($n = 24$). For all of these cases, the Risk management items were coded for the context outside. The HCR-20 was also coded for all patients who were newly admitted to the hospital during the period of this study ($n = 49$) and for inpatients at the request of their treatment team ($n = 17$), for instance, when they had questions about treatment progress. For these two types of patients, the Risk management items were coded for the context inside (risk of inpatient violence).

A researcher, group leader and treatment supervisor independently coded the HCR-20 for each case. When the patient was a sex offender, the Dutch version of the Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997; authorized Dutch version: Hildebrand, de Ruiter, & van Beek, 2001) was coded in addition to the HCR-20.[6] All raters had access to file information that consisted of psychological reports, reports to the court regarding treatment progress and recommendations for termination or prolongation of the tbs order, treatment plans and evaluations. Raters agreed upon a consensus score and a final risk judgment during a case conference. In these case conferences, raters also discussed the possibility of additional risk factors, protective factors and risk management strategies. The case conferences lasted on average about 1 hour and were considered useful by both researchers and clinicians. The results of the consensus meetings were used by staff to develop risk management strategies or for decision-making regarding leave or entry into the transmural treatment phase. In this sense, the HCR-20 judgments were used in the way that they are intended by the original developers of the instrument.

Initially, there were HCR-20 ratings of 149 patients; 127 men and 22 women. In a previous study, which included all women from the present study, we examined differences between 42 male and 42 matched female patients in mean scores and predictive validity of the HCR-20 (see de Vogel & de Ruiter, 2005).[7] Besides several differences in sample characteristics and mean individual HCR-20 item scores, we found that, except for the final risk judgment, the HCR-20 did not significantly predict violent recidivism in women as it did in men. Therefore, we decided to exclude the female patients from the present study. During the time course of this study, three (2%) patients had died (two by suicide, one by natural death). We decided not to exclude these patients because they all had a reasonably long follow-up period (19, 17 and 13 months, respectively). Also, during the time course of

this study, 20 (16%) patients were discharged from the hospital; 19 because their tbs order had been terminated by the court and one patient was readmitted to another forensic psychiatric hospital. We did not possess information on violent recidivism after termination of the tbs order. The mean follow-up period in treatment was 29.8 months (SD = 8.3, range = 9–37) for this group of 20 patients. We considered this follow-up period reasonably long and decided not to exclude these patients from the analyses. Furthermore, it should be noted that 19 (15%) patients were assessed more than once, because their leave situation had changed, for instance, they started with transmural treatment. The most recent risk assessment was used.

HCR-20 scores and final risk judgments were related to incidents of physical violence during treatment that occurred after the date of the most recent risk assessment (see *Violent outcome data*). Violent outcome data were collected until 1 June 2004. The mean follow-up period of the 127 patients was 21.5 months (SD = 10.9, range = 1–37).

Raters

The researchers ($n = 9$) were all Master's level clinical psychologists at the Research department, and were responsible for psychological assessment and empirical research in the hospital. The researchers were not in a treatment relationship with patients and did not have intensive contact with them, but they all knew the patients superficially. The treatment supervisors ($n = 8$) had a supervising and planning role in the treatment of around 20 patients; they were all senior clinicians, mostly clinical psychologists or psychotherapists. The professional background of the group leaders ($n = 59$) varied, but most of them had relevant higher vocational or academic training (e.g. nursing, social work, psychology). Group leaders conducted the daily and practical supervision and spent most of their time with the patients.

Violent outcome data

To identify incidents of physical violence, 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). Violent outcome data 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 last 24 hours are reported, such as disruptive incidents or positive results on urine analysis to detect if a patient has taken drugs. Incidents could have occurred inside the hospital (inpatient violence) or outside the hospital, for instance, for patients who were in the transmural treatment phase. We did not obtain data on violent recidivism after termination of the tbs order from the Ministry of Justice, because it was a rather small group whose tbs order had been terminated by the court ($n = 20$) and their mean follow-up period after discharge was quite brief (15 months, SD = 8.8, range = 4–34) compared to their mean follow-up period in (transmural) treatment (29.8 months, SD = 8.3, range = 9–37). Disruptive incidents were registered by the first author and assigned to one of four categories: verbal abuse, verbal threat, physical violence, and violation of hospital rules (see Hildebrand, de Ruiter, & Nijman, 2004 for details of the coding system). In this study, we focused on physical violence, more specifically on incidents of physical violence directed towards other persons, because the HCR-20 is designed to assess risk of violence to others. For instance, property damage alone was not included, unless the property damage occurred with the goal to frighten or threaten another person (e.g. smashing a cup of hot coffee against the wall while someone is standing close by). In order to examine if the HCR-

20 is able to predict different types of violence, we also considered verbal abuse and verbal threat.

Statistical analyses

The *F*-test was used to examine differences between researchers, group leaders and treatment supervisors on HCR-20 subscales and total scores. For differences in HCR-20 final risk judgments, chi-square tests were used. The predictive validity was established with Receiver Operating Characteristics (ROC) analyses (see for a detailed description Douglas, Guy & Weir 2005; Mossman, 1994; Rice & Harris, 1995). 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 (1 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 to large, and values above 0.75 as large (Douglas, Guy & Weir 2005). To compare the AUC values for the HCR-20 ratings of the three rater groups and the consensus, 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 and survival analyses, i.e. Cox regression (event history analyses) and Kaplan–Meier (see Tabachnick & Fidell, 2001) were conducted for comparative purposes. Survival analyses control for unequal follow-up periods between patients. Cox regression analyses, which result in the Hazard ratio (e^B) that can be interpreted as the relative risk, were conducted to determine which HCR-20 items were significant predictors. Cox regression analyses were also conducted to evaluate whether the HCR-20 final risk judgment added incremental validity to the HCR-20 actuarial scores. All analyses were conducted using SPSS version 11.

Results

Violent outcome

Nineteen patients (15%) committed a total of 27 incidents of physical violence during the period of this study, of which 14 committed one, three committed two, one committed three and one four. Accounting for time the patients had been at risk and using survival analysis, the failure rate was 23%. Examples of violent incidents were hitting another patient, attacking a staff member and throwing a table towards a window behind which staff members were standing. Two incidents occurred outside the hospital, the rest inside the hospital.[8] Most of the incidents of physical violence (82%) were classified as mildly serious and 18% as serious. In 63% of the incidents, staff members were the victim, in 30% other patients and in 7% the patient's girlfriend. Furthermore, 47 (37%) patients were registered for incidents of verbal abuse and 24 (20%) for incidents of verbal threat.

Risk assessments

Table I presents mean HCR-20 scores of the patients as coded by the three different rater groups, as well as mean HCR-20 scores as agreed upon in case conferences. Group leaders,

Table I. Risk assessments ($n = 127$).

	HCR-20 mean scores (SD)				HCR-20 final risk judgments		
	H scale	C scale	R scale	Total	Low	Moderate	High
Researchers	14.5 (3.1)	5.3 (2.1)	6.3 (2.2)[a]	26.1 (6.1)[a]	24%	45%	31%
Treatment supervisors	14.3 (3.4)	5.3 (2.2)	6.2 (2.2)[a]	25.8 (6.1)[a]	30%	46%	24%
Group leaders	14.0 (3.4)	5.0 (2.0)	5.3 (2.2)[b]	24.1 (5.8)[b]	21%	43%	35%
Consensus	14.8 (3.1)	5.5 (2.1)	6.4 (1.9)[a]	26.8 (5.6)[a]	28%	48%	24%

Note. [a > b], $p < 0.05$ (two-tailed). H scale = Historical scale. C scale = Clinical scale. R scale = Risk management scale. SD = standard deviation.

compared to researchers and treatment supervisors, gave significantly lower scores on the Risk management items and HCR-20 total score. There were no significant differences in mean HCR-20 scores between researchers and treatment supervisors. The mean HCR-20 consensus scores were higher – although not significantly – than the mean HCR-20 scores of the three individual rater groups. Table I also presents the HCR-20 final risk judgments. There were no significant differences between the rater groups in final risk judgments.

Predictive validity of the HCR-20 consensus ratings

AUC values and Pearson correlations for the HCR-20 subscales, total scores and final risk judgments as agreed upon by the three raters in case conferences were highly significant for incidents of physical violence during treatment (see Table II and Figure 1).

Patients who scored above the HCR-20 median of 27 compared to those who scored below had significantly more incidents of physical violence (failure rates as computed with Kaplan–Meier analysis: 2 versus 43, log rank (1, $n = 127$) = 15.8, $p < 0.001$; Odds ratio = 21.6, 95% CI = 2.8–167.2). The difference in failure rates between patients who were judged to be low, moderate or high risk was also significant (respectively 0, 8, and 64, log rank (2, $n = 127$) = 34.9, $p < 0.001$). Next, we conducted Cox regression analyses. The HCR-20 subscale scores were entered on block 1. The HCR-20 final risk judgment was entered on block 2 using the forward conditional method. In block 1, the HCR-20 subscales scores produced a significant model fit (χ^2 (3, $n = 127$) = 22.9, $p < 0.001$). In block 2, the HCR-20 final risk judgment produced a significant improvement to the model's fit (χ^2 change (1, $n = 127$) = 6.8, $p < 0.01$).

Table II. Predictive validity of the HCR-20 for physical violence ($n = 127$).

	Consensus			Researchers			Treatment supervisors			Group leaders		
	AUC	SE	*r*	AUC	SE	*r*	AUC	SE	*r*	AUC	SE	*r*
Historical scale	0.77***	0.05	0.32**	0.73***	0.06	0.27**	0.74***	0.06	0.28**	0.75***	0.06	0.29**
Clinical scale	0.80***	0.05	0.36**	0.76***	0.06	0.31**	0.75***	0.05	0.31**	0.66*	0.06	0.19*
Risk management scale	0.79***	0.05	0.35**	0.74***	0.06	0.29**	0.71***	0.05	0.27**	0.63	0.07	0.16
Total score	0.85***	0.04	0.43**	0.79***	0.05	0.35**	0.81***	0.05	0.36**	0.75***	0.05	0.30**
Final risk judgment	0.86***	0.04	0.49**	0.77***	0.06	0.35**	0.75***	0.05	0.33**	0.64*	0.07	0.19*

Note. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (two-tailed). AUC = Area Under the Curve. SE = Standard Error. r = Pearson point-biserial correlation.

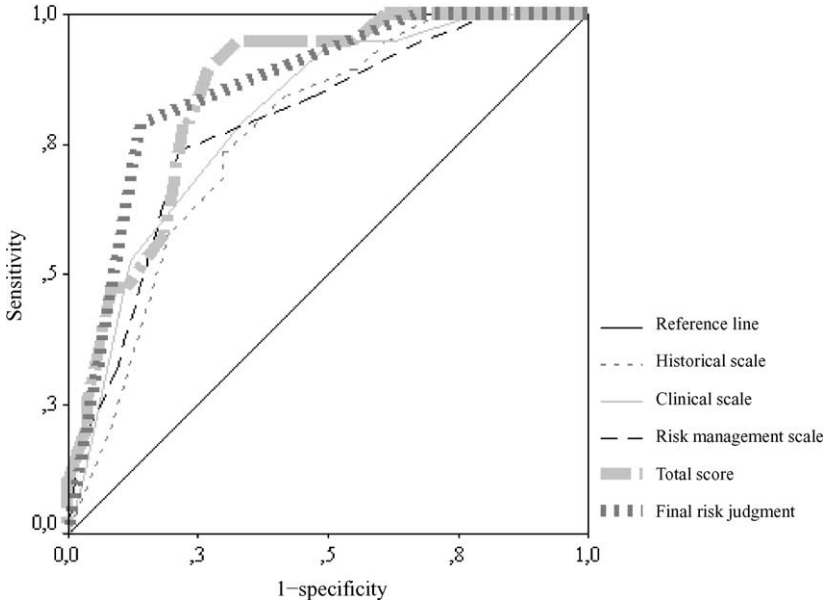

Figure 1. ROC curve HCR-20 consensus ratings for physical violence ($n = 127$).

Furthermore, we wanted to examine how the individual HCR-20 items perform in predicting physical violence in our sample. Table III shows the AUC values[9] and Pearson correlations for the consensus HCR-20 item scores and violent incidents. Items 2, 4, 5, 7, 11, 12, 14, 15, 16, 17, and 19 had significant AUC values and correlations. When Cox regression analyses were conducted, the full model with all HCR-20 items was found to be significant (χ^2 (20, $n = 127$) = 43.7, $p < 0.01$). Next, the forward conditional method was used to determine which HCR-20 items were significant predictors of incidents of physical violence. In the final model, items 2 ($e^B = 6.4$, 95% CI = 1.5–28.0), 15 ($e^B = 3.4$, 95% CI = 1.5–8.1), and 17 ($e^B = 3.4$, 95% CI = 1.2–10.0) were significant predictors of incidents of physical violence.

Although not displayed in the tables and figure, we also computed AUC values and Pearson correlations for the HCR-20 consensus ratings with respect to incidents of verbal abuse and verbal threat. We found significant predictive accuracy of the HCR-20 for both verbal abuse (total score: AUC = 0.72, SE = 0.05, $r = 0.36$, $p < 0.01$; final risk judgment: AUC = 0.65, SE = 0.05, $r = 0.28$, $p < 0.01$) and verbal threat (total score: AUC = 0.79, SE = 0.05, $r = 0.36$, $p < 0.01$; final risk judgment: AUC = 0.71, SE = 0.05, $r = 0.31$, $p < 0.01$).

Differences between raters in accuracy of risk assessments

AUC values and Pearson correlations for the HCR-20 ratings of the three rater groups were significant for incidents of physical violence (see Table II). One exception is the AUC value and correlation for the Risk management scale coded by the group leaders. With AccuROC we computed if there were significant differences in AUC values between the three rater groups and between individual group and consensus ratings. Group leaders compared to researchers had a significantly lower AUC value for the final risk judgment (χ^2 (1, $n = 127$) = 6.3, $p < 0.01$). Group leaders' ratings compared to consensus ratings also had significantly lower AUC values for the Clinical and Risk management scales, total score and

Table III. Predictive validity of the HCR-20 consensus items for physical violence ($n = 127$).

		AUC	SE	*r*
Historical items				
1.	Previous violence	0.48	0.07	−0.12
2.	Young age at first violent incident	0.72**	0.06	0.32**
3.	Relationship instability	0.60	0.06	0.16
4.	Employment problems	0.65*	0.06	0.21*
5.	Substance use problems	0.67*	0.06	0.24**
6.	Major mental illness	0.52	0.06	0.01
7.	Psychopathy	0.71**	0.06	0.29**
8.	Early maladjustment	0.57	0.07	0.11
9.	Personality disorder	0.53	0.07	0.08
10.	Prior supervision failure	0.58	0.06	0.15
Clinical items				
11.	Lack of insight	0.70**	0.06	0.27**
12.	Negative attitudes	0.71**	0.07	0.27**
13.	Active symptoms of major mental illness	0.47	0.05	−0.05
14.	Impulsivity	0.72**	0.06	0.29**
15.	Unresponsive to treatment	0.73***	0.07	0.32**
Risk management items				
16.	Plans lack feasibility	0.69**	0.07	0.26**
17.	Exposure to destabilizers	0.74***	0.06	0.33**
18.	Lack of personal support	0.61	0.07	0.16
19.	Noncompliance with remediation attempts	0.67*	0.06	0.25**
20.	Stress	0.57	0.07	0.13

Note. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (two-tailed). AUC = Area Under the Curve. SE = Standard Error. *r* = Pearson point-biserial correlation.

final risk judgment (χ^2 (1, $n = 127$) = respectively 6.8, 4.9, 4.6 and 20.1, $p < 0.05$). The AUC value for the HCR-20 consensus final risk judgment was significantly higher than the individual final risk judgment of researchers, treatment supervisors and group leaders (χ^2 (1, $n = 127$) = respectively 6.9, 5.3 and 20.1, $p < 0.01$).

Discussion

This is the first prospective study into the predictive validity of the Dutch HCR-20 in forensic clinical practice with multiple raters, including treating clinicians. The results of this study, which explored differences in predictive accuracy between researchers and clinicians and between individual and consensus ratings, provide strong support for the SPJ model of risk assessment.

The base rate of incidents of physical violence during treatment in this study was rather low compared to other studies (Dernevik et al., 2002; Nicholls, 2001; Ross, Hart, & Webster, 1998). However comparison of base rates of (inpatient) violence from different studies is complicated because of differences in samples, settings, length of follow-up periods, and definitions of violence. The low base rate of physical violence during treatment is possibly due to the structured, restrictive environment in which patients live inside a secure hospital. Another explanation might be the use of adequate risk management strategies by staff (e.g. isolating a patient when violence is expected).

We found good predictive validity of the HCR-20 for incidents of physical violence during treatment. This resembles findings from previous studies (Belfrage et al., 2000; Dernevik

et al., 2002; Gray et al., 2003; Ross et al., 1998). The AUC values for the HCR-20 consensus total scores and final risk judgment are quite high (AUC = 0.85 and 0.86, respectively) compared to those found in previous studies in forensic psychiatric samples (AUC values for HCR-20 total scores in other studies ranged from 0.57 to 0.84; see Douglas et al., 2005). A possible explanation is that our study was prospective and conducted in actual clinical practice implying that all raters personally knew the patients, had access to comprehensive file information and also had the opportunity to observe and monitor patients. This is in contrast to file based studies in which researchers retrospectively coded the HCR-20 from file information and did not know the subjects. The final risk judgment added significant incremental validity to the HCR-20 subscales scores, a finding similar to Douglas et al.'s (2003). Furthermore, we found that in our sample the consensus HCR-20 ratings were also predictive of verbal abuse and verbal threat. This finding is in accordance with a recently conducted study in 34 mentally disordered offenders that found the HCR-20 to be predictive of both verbal and physical aggression, but not of self-harm (Gray et al., 2003). With respect to the predictive accuracy of the HCR-20 items and subscales as agreed upon by the three raters in case conferences, the three subscales were found to have comparable predictive accuracy. However, although the differences between the AUC values for the subscales were small, the AUC values for the individual items show that several of the Historical items were not predictive in our sample, whereas most of the dynamic items were. The same pattern was found in two Swedish studies with high risk samples (Belfrage et al., 2000; Strand et al., 1999). A possible explanation is that most of the Historical risk factors are highly prevalent in high risk samples and thus do not discriminate between cases. In our sample, 125 (98%) patients had a score of 2 on Previous violence and 114 (90%) had a score of 2 on Personality disorder. Psychopathy was one of the Historical items that did demonstrate significant predictive accuracy, however. We want to emphasize this, because in our experience some mental health professionals decide to omit the Psychopathy item because the administration of the PCL-R is time-consuming and requires trained raters (see also Webster et al., 2002). Our finding underlines the statement of Hart (1998) that "psychopathy is a factor that should be considered in any assessment of violence risk" (p. 368). Most of the dynamic items were significant in predicting incidents of physical violence during treatment. In a recently conducted study in 100 psychiatric patients, the Clinical subscale was found to be specifically predictive of inpatient violence in the short term, whereas the Historical subscale was not (McNiel, Gregory, Lam, Binder, & Sullivan, 2003). Dynamic items that were not predictive in our sample were Active symptoms of major mental illness (almost absent: 106 (84%) patients had a score of 0 on this item), Lack of personal support and Stress (highly prevalent: 105 (83%) patients had a score of 2 on this item). Items that were most predictive (i.e. remained significant predictors in stepwise Cox regression analyses) were Young age at first violent incident, Unresponsive to treatment, and Exposure to destabilizers.

Next, differences were explored between the three rater groups in HCR-20 ratings and predictive accuracy. The group leaders compared to the other two groups gave significantly lower scores on the Risk management items. Regarding the HCR-20 mean scores and final risk judgments, no differences were found between researchers and treatment supervisors. In our previous study with 60 patients, we found that treatment supervisors compared to researchers significantly more often judged patients as low risk. In the present study, this difference was no longer significant. It should be noted that the results of these first 60 risk assessments were presented to treatment staff. Possibly, the treatment supervisors were influenced by the results and changed their way of rating the HCR-20. It is worth noting

that the consensus scores by the three raters were higher – although not significantly – than the scores of the three individual rater groups. This did not seem to affect the final risk judgment, however. The trend that group ratings lead to higher scores than individual ratings has been found before. For instance, Logan and Watt (2001) found that group ratings of the SVR-20 in 32 sex offenders were higher than individual ratings.

Regarding the predictive accuracy, not many differences between researchers and clinicians were found. When we started this research project, we expected differences between the researchers and the clinicians (group leaders and treatment supervisors), because of their different roles in the forensic setting. There was no difference, however, between researchers and treatment supervisors in accuracy of their risk assessments. In the previous study, feelings of clinicians towards their patients were found to be associated with the risk assessments, for instance, the feeling of being controlled and manipulated by the patient was related to higher HCR-20 scores (de Vogel and de Ruiter, 2004). The present findings suggest that treatment supervisors' feelings towards their patients did not interfere with the accuracy of their risk assessments. The finding that experienced clinicians were as accurate in using the HCR-20 as researchers, who were much more used to using structured instruments, is important because the HCR-20 is intended to be used by clinicians in their daily practice. The group leaders compared to the other two groups performed worse in predicting violence with the HCR-20. We offer three possible explanations. First, there was a large number of group leaders ($n = 59$) who participated in this study. Many group leaders conducted only one ($n = 22$), two ($n = 13$), or three ($n = 11$) risk assessments. Thus, group leaders compared to researchers and treatment supervisors gained less experience in coding the HCR-20. Second, group leaders compared to treatment supervisors and researchers were younger and less experienced in the forensic field. A third explanation is related to their role in treatment and their proximity to patients. Possibly, the group leaders' feelings towards their patients did interfere with their ability to objectively assess the risk of violence. For instance, several group leaders indicated that they found it difficult to be objective about a patient when they had just experienced an emotional outburst of this patient (see also Chakhssi & Hilterman, 2004).

Interestingly, the consensus risk assessments performed better than the risk assessments of the individual rater groups. This is especially true for the consensus final risk judgment, which was significantly better than the judgment of the three rater groups individually. Thus, ratings based on elaborate discussion with colleagues are superior to individual ratings. To our knowledge, no studies have been published before that compare HCR-20 consensus ratings to individual ratings. However, our finding is in line with previous studies that found higher predictive validity of clinical violence risk assessments when there was agreement between clinicians (McNiel et al., 2000) or when clinicians' ratings were aggregated as "group" ratings (Huss & Zeiss, 2004). In conclusion, the findings demonstrate that the method of SPJ, i.e. systematically rating risk factors, integrating and weighing information to arrive upon a final risk judgment and discussion with colleagues, is effective in predicting future violence risk.

There are several limitations to this study. First, prospective predictive research is hampered by the clinical goals of risk assessment, i.e. risk management and prevention (Dernevik et al., 2002; Hart, 1998). Hart (1998) stated that predictions of violence are not passive assessments, but decisions that influence services delivered to individuals: "Clinicians are bound – morally, ethically, and legally – to try to prove themselves wrong when they predict violence and take every reasonable action to prevent violence" (p. 365). In our study, clinicians were able to use the results of the HCR-20 ratings, for instance, for

decisions concerning leave. Thus, it is very likely that risk management was influenced by the results of the risk assessment, for instance, high risk patients were not released from the hospital, or were separated if the risk of inpatient violence was judged to be high. So, the AUC values we obtained were already high, but might have been even higher if the results had not been used to manage risk. Second, the sample was derived from only one Dutch forensic psychiatric hospital, thereby limiting generalization. Nevertheless, we consider this group to be representative of Dutch offenders with a tbs order, because they are largely similar in demographic, psychiatric and criminal characteristics to the total population of tbs offenders (see van Emmerik & Brouwers, 2001). Third, the mean follow-up period of this study was somewhat limited; some patients had a very short follow-up period of only 1 or 2 months. Also, the range of follow-up periods was rather large (1–37 months), which complicates comparison between patients. Nevertheless, the survival analyses we conducted take differences in time-at-risk into account. Fourth, we found a rather low base rate of violence. Although we conducted ROC analyses that are insensitive to base rates, the low base rate might have had an effect on the Cox regression analyses. A final limitation is that data regarding violent outcome were not always reliable. 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 told to staff. This is the case for inpatients, but even more so for patients who are in the transmural treatment phase or who can go outside the hospital without supervision. It should be noted, however, that most of these limitations would have had a negative effect on the predictive accuracy of the HCR-20, thus the findings might have been even stronger without these limitations.

Based on our findings and experiences we would like to conclude with some recommendations for the use of the HCR-20 in forensic clinical practice. Although it is clearly stated in the HCR-20 manual (Webster et al., 1997) and still more recently pointed out by Webster et al. (2002), we want to emphasize again that raters should be trained and experienced in performing risk assessments with the HCR-20. In addition, they should keep their skills up to date by advanced training, keeping up with the literature on violence risk assessment, and performing risk assessments on a regular basis. In this study, there was a requirement for group leaders to conduct at least one risk assessment every 6 months or they had to repeat the training. Furthermore, we strongly recommend to have more than one rater coding the HCR-20, preferably raters with different roles in treatment, for instance, an objective, more distant person like a researcher or diagnostician, and an experienced clinician who knows the patient well. Structured discussion about the risk factors in a case conference is very useful and can improve the accuracy of the risk assessment. Moreover, it can help to design risk management strategies. The identification of possible protective factors is important because the aim is to minimize violence risk. Finally, it is important to repeat the violence risk assessment every time the context changes, for instance, when the liberties of patients are expanded.

Acknowledgements

The authors thank all clinicians and researchers who participated in this study. Special thanks go to Cécile Vandeputte-van de Vijver who functioned as workshop trainer together with the first author and participated as a researcher in the study.

Notes

1 See Douglas, Cox, and Webster (1999) and Otto (2000) for a more detailed overview of risk assessment approaches.
2 This sample comprised 53 men and seven women. In the present study, the women were excluded from the analyses (see *Procedure*).
3 In Dutch forensic psychiatry, cluster B personality disorders are the most prevalent (see Hildebrand & de Ruiter, 2004; de Ruiter & Greeven, 2000).
4 Axis I diagnoses were lifetime clinical diagnoses based on consensus between four raters (see Hildebrand & de Ruiter, 2004), Axis II disorders were diagnosed with the Structured Interview for DSM-IV Personality (SIDP-IV; Pfohl, Blum, & Zimmerman, 1995).
5 This 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.
6 Results of the SVR-20 are not included in this study but can be expected within 1 or 2 years.
7 The HCR-20 was developed based on research in predominantly male samples.
8 Five of seven patients that were assessed for the context Out had an incident of physical violence inside the hospital. Four of these patients were in the transmural phase. Although these patients lived outside the hospital, they frequently visited the hospital, for instance, to attend work or psychotherapy.
9 It should be noted that ROC analyses are less appropriate to apply with dichotomous or trichotomous variables. Still, we believed it was important to examine the predictive accuracy per item, and AUC values are easy to understand and provide comparison values with other similar studies.

References

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). Washington, DC: APA.

Belfrage, H. (1998). Implementing the HCR-20 scheme for risk assessment in a forensic psychiatric hospital: Integrating research and clinical practice. *Journal of Forensic Psychiatry*, *9*, 328–338.

Belfrage, H., Fransson, G., & Strand, S. (2000). Prediction of violence using the HCR-20: A prospective study in two maximum-security correctional institutions. *Journal of Forensic Psychiatry*, *11*, 167–175.

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk-20. Professional guidelines for assessing risk of sexual violence*. Vancouver, BC: British Columbia Institute against Family Violence.

Chakhssi, F., & Hilterman, E. (2004). Psychiatric staff's perceptions and judgment of violence risk. Paper presented at the Fourth Annual Conference of the International Association of Forensic Mental Health Services, Stockholm, Sweden.

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. *Journal of Forensic Psychiatry and Psychology*, *15*, 145–164.

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 and Psychotherap*, *12*, 226–240.

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 treated 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 nonparametric approach. *Biometrics*, *44*, 837–854.

Dernevik, M. (2004). *Structured clinical assessment and management of risk of violent recidivism in mentally disordered offenders*. Stockholm, Sweden: Karolinska University Press.

Dernevik, M., Falkheim, M., Holmqvist, R., & Sandell, R. (2001). Implementing risk assessment procedures in a forensic psychiatric setting: Clinical judgement revisited. In D. P. Farrington, C. R. Hollin, & M. McMurran (Eds.), *Sex and violence: The psychology of crime and risk assessment* (pp. 83–101). London: Routledge.

Dernevik, M., Grann, M., & Johansson, S. (2002). Violent behaviour in forensic psychiatric patients: Risk assessment and different risk-management levels using the HCR-20. *Psychology, Crime & Law*, *8*, 93–111.

Dolan, M., & Khawala, A. (2004). The HCR-20 and post-discharge outcome in male patients discharged from medium security in the UK. *Aggressive Behavior*, *30*, 469–483.

Douglas, K. S., Cox, D. N., & Webster, C. D. (1999). Violence risk assessment: Science and practice. *Legal and Criminological Psychology*, *4*, 149–184.

Douglas, K. S., Guy, L. S., & Weir, J. (2005). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*. Available online at: http://www.sfu.ca/psychology/groups/faculty/hart/violink.htm.

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., Ogloff, J. R. P., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology*, *67*, 917–930.

Gray, N. S., Hill, C., McGleish, A., Timmons, D., MacCulloch, M. J., & Snowden, R. J. (2003). Prediction of violence and self-harm in mentally disordered offenders: A prospective study of the efficacy of HCR-20, PCL-R and psychiatric symptomatology. *Journal of Consulting and Clinical Psychology*, *71*, 443–451.

Hanson, R. K., & Thornton, D. (2002). *Notes on the development of the Static-2002*. Ottawa: Public Works and Government Services Canada.

Hare, R. D. (1991). *Manual for the Hare Psychopathy Checklist – Revised*. Toronto, Ontario: Multi-Health Systems.

Hare, R. D. (2003). *Hare Psychopathy Checklist-Revised Second Edition. Technical Manual*. Toronto, Ontario: Multi-Health Systems.

Harris, G. T., & Rice, M. E. (1997). Risk appraisal and management of violent behavior. *Psychiatric Services*, *48*, 1166–1176.

Hart, S. D. (1998). Psychopathy and the risk for violence. In D. J. Cooke, A. E. Forth, & R. D. Hare (Eds.), *Psychopathy: Theory, research and implications for society* (pp. 355–373). Dordrecht: Kluwer.

Hart, S. D., Cox, D., & Hare, R. D. (1995). *The Hare Psychopathy Checklist: Screening Version (PCL:SV)*. Toronto, Ontario, Canada: Multi-Health Systems.

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., & 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.

Hildebrand, M., de Ruiter, C., & van Beek, D. (2001). *SVR-20. Richtlijnen voor het beoordelen van het risico van seksueel gewelddadig gedrag* [SVR-20. Guidelines for the assessment of risk of sexual violence]. Utrecht, The Netherlands: Forum Educatief.

Huss, M. T., & Zeiss, R. A. (2004). Clinical assessment of violence from inpatient records: A comparison of individual and aggregate decision making across risk strategies. *International Journal of Forensic Mental Health*, *3*, 139–147.

Laws, D. R., Hudson, S. M., & Ward, T. (2000). *Remaking relapse prevention: A sourcebook*. London: Sage.

Litwack, T. R., & Schlesinger, L. B. (1999). Dangerousness risk assessments: Research, legal, and clinical considerations. In A. K Hess, & I. B. Weiner (Eds.), *The handbook of forensic psychology* (pp. 171–217). New York: Wiley.

Logan, C., & Watt, K. (2001). Structured professional guidelines approaches to risk assessment: Single practitioner vs. multidisciplinary team administration. Paper presented at the International Conference Violence Risk Assessment and Management: Bridging science and practice closing together, Sundsvall, Sweden.

McNiel, D., Gregory, A., Lam, J., Binder, R., & Sullivan, G. (2003). Utility of decision support tools for assessing acute risk of violence. *Journal of Consulting and Clinical Psychology*, *71*, 945–953.

McNiel, D., Lam, J., & Binder, R. (2000). Relevance of interrater agreement to violence risk assessment. *Journal of Consulting and Clinical Psychology*, *68*, 1111–1115.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, *62*, 783–792.

Nicholls, T. L. (2001). Violence risk assessment with female NCRMD acquittees: Validity of the HCR-20 and PCL:SV. Unpublished master's thesis, Simon Fraser University, Vancouver, British Columbia, Canada.

Otto, R. K. (2000). Assessing and managing violence risk in outpatient settings. *Journal of Clinical Psychology*, *56*, 1239–1262.

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 gewelddadig gedrag. Versie 2* [HCR-20. Assessing the risk of violence. Version 2]. Nijmegen/Utrecht, The Netherlands: Prof. Mr. W. P. J. Pompestichting/Dr. Henri van der Hoeven Stichting.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, *63*, 737–748.

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.

Strand, S., Belfrage, H., Fransson, G., & Levander, S. (1999). Clinical and risk management factors in risk prediction of mentally disordered offenders: More important than actuarial data? *Legal and Criminological Psychology*, *4*, 67–76.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.

van Emmerik, J. L., & Brouwers, M. (2001). *De terbeschikkingstelling in maat en getal: Een beschrijving van de tbs-populatie in de periode 1995–2000* [The tbs-order in numbers and figures. A description of the tbs population in the period 1995–2000]. Ministerie van Justitie, The Hague: Dienst Justitiële Inrichtingen.

Vida, S. (1997). *AccuROC nonparametric receiver operating characteristic analysis* (version 2.5) [Computer software]. Montreal, Quebec, Canada: Accumetric Corporation.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20. Assessing the risk of violence. Version 2*. Vancouver, BC, Canada: Simon Fraser University and Forensic Psychiatric Services Commission of British Columbia.

Webster, C. D., Müller-Isberner, R., & Fransson, G. (2002). Violence risk assessment: Using structured clinical guidelines professionally. *International Journal of Forensic Mental Health*, *1*, 185–193.