

Predictive Validity of the SVR-20 and Static-99 in a Dutch Sample of Treated Sex Offenders

Vivienne de Vogel,^{1,5} Corine de Ruiter,^{2,3} Daan van Beek,¹ and Gwen Mead⁴

In this retrospective study, the interrater reliability and predictive validity of 2 risk assessment instruments for sexual violence are presented. The SVR-20, an instrument for structured professional judgment, and the Static-99, an actuarial risk assessment instrument, were coded from file information of 122 sex offenders who were admitted to a Dutch forensic psychiatric hospital between 1974 and 1996 (average follow-up period 140 months). Recidivism data (reconvictions) from the Ministry of Justice were related to the risk assessments. The base rate for sexual recidivism was 39%, for nonsexual violent offenses 46%, and for general offenses 74%. Predictive validity of the SVR-20 was good (total score: $r = .50$, $AUC = .80$; final risk judgment: $r = .60$, $AUC = .83$), of the Static-99 moderate (total score: $r = .38$, $AUC = .71$; risk category: $r = .30$, $AUC = .66$). The SVR-20 final risk judgment was a significantly better predictor of sexual recidivism than the Static-99 risk category.

KEY WORDS: risk assessment; sexual recidivism.

The assessment of risk for (sexual) violence is an important task of psychologists working in forensic practice. Sexual violent (re)offending often has severe consequences for the victims and causes strong feelings of fear, anger, and concern in society. A carefully conducted risk assessment before a probationary leave, parole decision, or termination of (mandatory) treatment can help to appraise the risk of recidivism in an adequate way and thereby prevent serious (sexual) violent offenses (Douglas & Webster, 1999). To date, the best known and most widely used method in practice, at least in the Netherlands, is the unstructured clinical judgment approach that is exclusively on the basis of professional expertise of the clinician. However, research has revealed some important limitations of this unstructured clinical judgment, such as poor reliability and validity (Monahan, 1981; see for a discussion of

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

²University of Amsterdam, Amsterdam, The Netherlands.

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

⁴Leiden University, Leiden, The Netherlands.

⁵To whom correspondence should be addressed at Department of Research, Dr. Henri van der Hoeven Kliniek, P.O. Box 174, 3500 AD Utrecht, The Netherlands; e-mail: vdevogel@hoevenstichting.nl.

these disadvantages (Quinsey, Harris, Rice, & Cormier, 1998, pp. 55–72) Although more recent studies have demonstrated clinical accuracy to be significantly better than chance, unstructured clinical judgment is liable to systematic biases. For example, clinicians were found to be accurate in predicting risk of recidivism in cases with a violent history, but less accurate in predicting risk of violence in female psychiatric patients (underestimation of risk) and nonwhite men (overestimation of risk; Lidz, Mulvey, & Gardner, 1993; McNiel & Binder, 1995).⁶ Therefore, several authors recommend to employ more structured risk assessment procedures in order to optimize accuracy and validity (Borum, 1996; Webster, Douglas, Eaves, & Hart, 1997a).

An important distinction among structured risk assessment instruments can be made between the actuarial and the structured professional judgment (SPJ) approaches. Actuarial instruments are developed on the basis of risk factors that are empirically related to (sexual) violent behavior. These instruments are relatively simple to code—according to fixed rules and not necessarily by a forensic expert—and contain predominantly static, nonchangeable factors that are added up according to a fixed algorithm to reach a conclusion on the risk of recidivism. Examples are the Violence Risk Appraisal Guide for violent behavior (VRAG; Harris & Rice, 1997) and the Static-99 for sexual violent behavior (Hanson & Thornton, 1999). Although risk assessment with actuarial instruments is a simple and time-effective procedure, there are some important disadvantages to this approach. Most of the actuarial instruments⁷ do not include situational or dynamic risk factors and do not offer guidelines for treatment, which makes them useless in treatment settings where the aim is reduction of the risk of recidivism. Furthermore, generalization towards populations other than the type of samples in which the instrument was developed is limited (Grubin & Wingate, 1996; Hart, 1998). On the basis of this criticism, a new risk assessment approach was developed, called structured professional judgment (SPJ). In this approach, the risk assessment is performed by a forensic clinician by means of a standardized checklist, containing empirically derived risk factors for (sexual) violence, historical 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. Examples of SPJ guidelines are the Historical, Clinical, Risk Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997b) and the Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997). Research in several populations and settings has demonstrated good interrater reliability and predictive validity of the above mentioned risk assessment instruments (e.g., Belfrage, Fransson, & Strand, 2000; Dempster, 1998; Douglas, Ogloff, Nicholls, & Grant, 1999). However, these results have been obtained with North American samples and some European samples, predominantly Swedish (see also the special issue of *Psychology, Crime and Law*; Hart, 2002); the psychometric properties of the Dutch translations of these risk assessment instruments are unknown.

⁶For a detailed discussion about the clinical-actuarial controversy, we refer the reader to two reviews: Douglas, Cox, and Webster (1999) and Litwack (2001).

⁷An exception is the Sex Offender Need Assessment Rating (SONAR) that completely consists of dynamic factors. Hanson and Harris (2000) developed this instrument as a supplement to the Static-99.

Sex offenders are considered as a special group for the assessment of risk of recidivism. Various studies and meta-analyses have indicated there are specific risk factors for sexual violence, aside from risk factors for general violence (e.g., psychopathy, criminal history), such as sexual deviance and prior sexual offenses (Hanson & Bussière, 1998; Hanson, Steffy, & Gauthier, 1993). Most of the risk assessment schemes for sexual violence were developed on the basis of findings from these studies. Furthermore, research has demonstrated that the group of sex offenders is heterogeneous one (Doren, 1998; Greenberg, 1998; Prentky, Lee, Knight, & Cerce, 1997), which makes it difficult to develop a risk assessment instrument that predicts accurately for all types of sex offenders. For example, there are major differences in the base rate for sexual reoffending between rapists, child molesters with extrafamilial boys as victim, child molesters with extrafamilial girls as victims, and incest offenders. Child molesters with extrafamilial boys as victims reoffend more than rapists and child molesters with extrafamilial girls as victims (Hanson et al., 1993; Quinsey, Lalumière, Rice, & Harris, 1995). Furthermore, some sex offenders, particularly child molesters, may reoffend after a long period of nonoffending (Hanson et al., 1993; Prentky et al., 1997). This makes it necessary to reassess the risk of recidivism regularly. When studying the literature on sexual recidivism, one can conclude that there are a number of subgroups of sex offenders that reoffend frequently and seriously. Risk assessment can assist in detecting these subgroups and distinguish them from sex offenders who pose a low or moderate risk of recidivism.

In this article, we will present findings from a retrospective study on the interrater reliability and predictive validity of two risk assessment instruments for sexual violence—the Static-99 and the SVR-20—in a group of sex offenders who were admitted to a Dutch forensic psychiatric hospital between 1974 and 1996. The aim of the present study was to determine the value of these instruments for the prediction of sexual violence in The Netherlands, and to compare the predictive value of the actuarial instrument with the guideline according to the SPJ approach.

METHOD

Setting

This study was conducted in the Dr. Henri van der Hoeven Kliniek, a 114-bed forensic psychiatric hospital in The Netherlands. Patients are admitted under the judicial measure *terbeschikkingstelling* (tbs) that can be translated as “disposal to be treated on behalf of the state.” The tbs-order is imposed by court on offenders who committed a serious offense and are considered to have diminished responsibility for it because of severe psychopathology. The tbs-order is of indefinite duration; every 1 or 2 years the court reevaluates 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 programs, for instance, job training, education, sports, creative arts, and psychotherapy. Since the 1980s, 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; van Beek, 1999). The emphasis of treatment is not on changing the personality of the offender, but on reducing/managing risk factors for recidivism. Family and friends of the patient are involved in the treatment, for instance, when the patient presents his relapse prevention plan.

Subjects

The group of sex offenders consisted of 95 rapists and 27 child molesters, all male. The child molester group was composed of 16 child molesters with extrafamilial girls as victim, 10 child molesters with extrafamilial boys as victim, and one incest offender. Table 1 presents the demographic characteristics of the sample. The majority of the sex offenders were Dutch, single, and had no work at the time of the index offense. More than half of the group did not complete their hospital treatment; in 36% of the cases, the tbs-order was terminated by court against the hospital’s advice, and 29% of the sex offenders were readmitted to another forensic psychiatric institution. Reasons for these replacements differed, but the most common was that the therapeutic

Table 1. Sample Characteristics

	Rapists <i>N</i> = 95	Child molesters <i>N</i> = 27	Total <i>N</i> = 122
Demographic			
Mean age upon admission	24.6	25.6	24.8
Dutch nationality	84 (88%)	25 (93%)	109 (89%)
Upbringing in foster or children’s home	48 (51%)	19 (70%)**	67 (55%)
Single (at the time of the index offense)	74 (78%)	25 (93%)	99 (82%)
No education after primary school	53 (56%)	14 (52%)	67 (55%)
Special education	19 (20%)	12 (44%)**	31 (25%)
Unemployed (at the time of the index offense)	49 (52%)	13 (48%)	62 (51%)
Psychiatric			
No psychiatric history	27 (28%)	2 (7%)*	29 (24%)
Outpatient treatment(s)	21 (22%)	4 (15%)	25 (21%)
Inpatient admission(s)	47 (50%)	21 (78%)**	68 (56%)
Alcohol abuse	34 (36%)	10 (37%)	44 (36%)
Drug abuse	3 (3%)	1 (4%)	4 (3%)
Multiple substance abuse	21 (22%)	3 (11%)	24 (20%)
Mean intelligence score	97.1 ^a	85.8 ^b	93.8
Offenses			
Victim was not a stranger	15 (16%)	10 (37%)*	25 (21%)
Number of victims more than one	35 (40%)	16 (59%)*	51 (42%)
Previously convicted for sex offense(s)	60 (63%)	20 (74%)	80 (66%)
Previously convicted to the tbs-order	5 (5%)	5 (19%)	10 (8%)
At the time of the study still or again under the tbs-order	4 (4%)	4 (15%)*	8 (7%)
Treatment			
Mean duration of treatment in months	51.4	62.9	54.0
Treatment included a probationary period	38 (40%)	11 (41%)	49 (40%)
Mean duration of probationary period in months	12.6	17.0	13.6
Readmitted to another institution	29 (31%)	6 (22%)	35 (29%)
Termination of tbs-order against the hospital’s advice	33 (35%)	11 (41%)	44 (36%)

Note. ^b < ^a, *p* < .01 (*F* = 7.6). Special education is for children with learning disabilities and/or conduct problems. Intelligence scores were available for 42 rapists and 17 child molesters.

p* < .05, *p* < .01 (chi-square analysis, two-tailed).

Table 2. Items of the Static-99

1. Prior sex offenses
2. Prior sentencing dates (excluding index offense)
3. Any convictions for noncontact sex offenses
4. Index nonsexual violence
5. Prior nonsexual violence
6. Any unrelated victims
7. Any stranger victims
8. Any male victim
9. Young (18–24 years)
10. Single (ever lived with lover for 2 years or more?)

Note. Adopted from Hanson and Thornton (1999).

relationship between the patient and hospital staff was disturbed to such an extent that further treatment was considered impossible. The table also shows that there were a number of significant differences between rapists and child molesters. Child molesters more often than rapists grew up in foster or children's homes, had been admitted to inpatient psychiatric hospitals, and obtained lower scores on intelligence scales. Furthermore, child molesters more often knew their victim and had made more than one victim compared to rapists.

Instruments

Static-99

The Static-99 is a brief actuarial instrument for the assessment of risk for sexual violence in adult sex offenders. The instrument is derived from a fusion of two previously developed risk assessment instruments, the Rapid Risk Assessment of Sexual Offense Recidivism (RRASOR; Hanson, 1997) and the Structured Anchored Clinical Judgment (SACJ-Min; Grubin, 1998). The Static-99 is composed of 10 historical risk factors (see Table 2) that have to be coded from file information. The factors add up to a maximum total score of 12 that is subsequently translated into four risk categories: low (0,1), medium low (2–3), medium high (4–5) and high (6 or more; Hanson & Thornton, 1999).

Hanson and Thornton (2000) tested the predictive accuracy of the Static-99 in a group of 1,301 sex offenders from four different institutions in Canada and England and found a moderate predictive validity for sexual violence ($r = .33$, Receiver Operating Characteristics [ROC]:⁸ Area Under the Curve [AUC] = .71) and violent recidivism ($r = .32$, AUC = .69). In this study, sexual offenses were included in the definition of violent recidivism. Similarly, Sjöstedt and Långström (2001) found moderate to good predictive validity in a group of 1,400 Swedish prisoners (sexual recidivism: $r = .22$, AUC = .76; violent (including sexual offenses) recidivism: $r = .30$, AUC = .74). This study rendered a good interrater reliability (Cohen's $\kappa = .90$).

⁸ROC is a statistical method to assess predictive validity. See also "Analyses."

Table 3. SVR-20 Items*Psychosocial adjustment*

1. Sexual deviance
2. Victim of child abuse
3. Psychopathy
4. Major mental illness
5. Substance use problems
6. Suicidal/homicidal ideation
7. Relationship problems
8. Employment problems
9. Past nonsexual violent offenses
10. Past nonviolent offenses
11. Past supervision failure

Sexual offenses

12. High density sex offenses
13. Multiple sex offense types
14. Physical harm to victim(s) in sex offenses
15. Uses weapons or threats of death in sex offenses
16. Escalation in frequency or severity of sex offenses
17. Extreme minimization or denial of sex offenses
18. Attitudes that support or condone sex offenses

Future plans

19. Lacks realistic plans
20. Negative attitude toward intervention

Note. Adopted from Boer, Hart, Kropp, and Webster (1997).

SVR-20

The SVR-20 is a structured clinical guideline (checklist) designed for the assessment of risk for sexual violence in adult sex offenders. The instrument was developed from a thorough consideration of the empirical literature and the clinical expertise of a number of clinicians. The SVR-20 consists of 20 items, divided into three domains: Psychosocial adjustment, Sexual offenses, and Future plans, that have to be coded by an experienced forensic clinician. The clinician should use all available information on the sex offender, preferably from different sources and gathered with different methods, for example, criminal records/police files, psychological reports, interviews with significant others, and observations. The items have to be coded on a 3-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*. Table 3 presents the items of the SVR-20. The Psychopathy Checklist-Revised (PCL-R; Hare, 1991)⁹ is used to code item 3 Psychopathy: a PCL-R score of 30 justifies a code 2 on the SVR-20, and a PCL-R score between 20 and 29 represents a code 1. Aside from the 20 items, the SVR-20 offers the possibility to code “other considerations,” that is, case-specific risk factors that do not fit within the descriptions of the 20 items.

The final risk judgment has to be indicated as low, moderate, or high and is valid for a specific time period and for a specific context (e.g., inpatient vs. outpatient). The key question for the low, moderate, or high judgment is: what level of effort, attention, and intervention is required to prevent this person from perpetrating sexual violence? The final risk judgment not only depends on the simple summation of item scores,

⁹For a more detailed description of the PCL-R, see below.

but also on specific combinations of factors or other considerations. For example, research revealed that the combination of psychopathy as measured by the PCL-R and sexual deviance, increases the risk of sexual recidivism considerably (Rice & Harris, 1997). In some cases, only one or two items may be sufficient to justify the judgment “high risk,” for example, when a sex offender has intrusive sadistic, homicidal thoughts and is known to have committed extremely violent rapes in the past. The final risk judgment can be considered as a SPJ that is arrived at through the process of coding the guideline and integrating all available information.

To date, little is known about the psychometric properties of the SVR-20. Dempster (1998, see also Dempster & Hart, 2002) examined the predictive validity of five risk assessment instruments: the SVR-20, PCL-R, VRAG, RRASOR, and the Sex Offender Risk Assessment Guide (SORAG; Quinsey et al., 1998). These instruments were rated for 95 sex offenders from several correctional institutions in Canada. All instruments predicted violence, but only the RRASOR and the SVR-20 final risk judgment were significant in predicting sexual violence. The clinical judgment of risk on the basis of SVR-20 had incremental predictive validity relative to the actuarial measures. Sjöstedt and Långström (2002) compared four risk assessment instruments, including the SVR-20 in a sample of 51 rapists and found a significant relation between the subscale Psychosocial adjustment and nonsexual violent recidivism, but no significant predictive validity of the SVR-20 for sexual offenses. However, the authors warrant caution in generalizing their results because this study had some important limitations such as a small sample size and poor interrater reliability for the SVR-20 (average Cohen's $\kappa = .36$).

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 3-point scale—“0” *item does not apply*, “1” *the item probably or partially applies*, and “2” *the item definitely applies*—from a semistructured 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 30, but in some countries, for instance, Scotland, England, and Sweden a cut off score of 25 or 26 has proven useful (Hare, Clark, Grann, & Thornton, 2000).

Several studies have demonstrated that the PCL-R total score is a strong predictor of general, sexual, and violent recidivism in both prison and general/forensic psychiatric populations (Hemphill, Templeman, Wong, & Hare, 1998; Salekin, Rogers, & Sewell, 1996). Hart (1998) stated that: “Psychopathy is a factor that should be considered in any assessment of violence risk” (p. 368). Therefore, psychopathy as measured by the PCL-R is included as one of the risk factors in most structured

risk assessment instruments, like the VRAG, HCR-20, and SVR-20. The ability of the PCL-R to predict recidivism was shown to have cross-cultural generalizability (Hare et al., 2000). Research in the Dr. Henri van der Hoeven Kliniek rendered a good interrater reliability for the Dutch version of the PCL-R (Vertommen, Verheul, de Ruiter, & Hildebrand, 2002): single measure Intraclass Correlation Coefficient (ICC) for PCL-R total score = .88, weighted average Cohen's kappa for the categorical diagnoses = .63 (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 in the Dr. Henri van der Hoeven Kliniek (Hildebrand, de Ruiter, & Nijman, 2004).

In the present study, the PCL-R was coded exclusively on file information. Ideally, the PCL-R is coded on the basis of both a semistructured interview and file information, however, previous research showed that for research purposes, PCL-R ratings can be done reliably on file information (Grann, Långström, Tengström, & Stålenheim, 1998; Hildebrand, de Ruiter, & de Vogel, 2004).

Procedure

File information was gathered on 123 sex offenders who were admitted between 1974 and 1996 in the hospital (release dates between 1977 and 2000). In general, these files consisted of psychological reports, reports to the court regarding treatment progress and recommendations for termination or prolongation, treatment plans, and evaluations. Next, we rated the Dutch versions of the SVR-20 (Hildebrand, de Ruiter, & van Beek, 2001), PCL-R (Vertommen et al., 2002), and Static-99 (van Beek, de Doncker, & de Ruiter, 2001) on the basis of all available file information. In order to establish the interrater reliability, three raters (in different compositions out of a group of four raters) independently rated 30 cases. During a case meeting, raters discussed their scores, and agreed upon a consensus score. The case meetings provided raters with an opportunity to sharpen their understanding of the individual SVR-20 items. The consensus score was used for the analyses on predictive validity. Subsequently, the remaining cases were divided among the four raters (VdV: 73, CdR: 1, DvB: 10, GM: 7). The rating procedure was conducted while all raters were blind to the outcome.¹⁰ One case was not rated, because this patient died within 2 months after admission, and one case was not included in the analyses because this patient was an illegal inhabitant of The Netherlands who immediately returned to his native country after termination of the tbs-order.

Recidivism Data

Data on recidivism were retrieved from the Judicial Documentation register of the Ministry of Justice after all the files had been coded. Sexual recidivism was defined as a new conviction for a sex offense in accordance with Dutch criminal law, and comprises both hands-on (e.g., rape, sexual assault, child molestation) and hands-off

¹⁰One of the four raters is a clinician who knew (about the outcome of) some of the patients. This clinician coded files of patients he did not know and the other files were randomly divided among the other raters who had no information about the outcome.

(e.g., exhibitionism, possession of child pornography) offenses. Furthermore, we explored new convictions for nonsexual violent and general offenses. The follow-up period, starting on the date of release from the hospital or readmission to another institution and ending on the date of data gathering (November 2001), varied from 20 to 291 months with an average of 140 months.

Analyses

F tests and chi-square tests were used to examine differences between the rapists and child molesters in demographic characteristics. Survival analyses, more specifically the Kaplan Meier method, were used to calculate recidivism rates (Schmidt & Wytte, 1988; Tabachnick & Fidell, 2001). These analyses take into account the time the offender has been at risk. Thus, it is possible to calculate the recidivism rate for a specific period despite the fact that the follow-up periods of the offenders diverge. The interrater reliability of the SVR-20 and Static-99 was examined by means of the ICC, using the two-way random effect variance model and consistency type (McGraw & Wong, 1996). Critical values used for single measure ICCs were: $ICC \geq .75$ = excellent; $.60 \leq ICC < .75$ = good; $.40 \leq ICC < .60$ = moderate; $ICC < .40$ = poor (Fleiss, 1986). The predictive validity of both the instruments was established with ROC analyses (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 AUC can be interpreted as the probability that a randomly selected recidivist would score higher on the instrument than a randomly selected nonrecidivist. An AUC of .50 represents chance prediction, and an AUC of 1.0 perfect prediction. In general, AUC values of .70 and above are considered moderate, and above .75 good (Douglas, 2001). To compare the obtained AUC values of the SVR-20 and Static-99, we used the software program AccuROC (Vida, 1997) which applies the nonparametric method described by DeLong, DeLong, and Clarke-Pearson (1988). Pearson *r* correlations were calculated for comparative purposes.

RESULTS

Reconviction Rates

The base rate for sexual recidivism was 39%, for nonsexual violent offenses 46%, and for general offenses 74% (rates computed with survival analyses: 48, 63, and 91, respectively). The Judicial registration system reported 89 new convictions for sexual offenses, 10 for homicide offenses, and 77 for nonsexual violent offenses. The most frequently reported sexual and violent offenses were: rape (25), sexual assault (12), exhibitionism (16), assault (19), threat (19), and unlawful confinement (16).

Interrater Reliability

The interrater reliability of the SVR-20 subscales and total score was good to excellent (ICC single measure: SVR-20 total score = .75, Psychosocial adjustment = .74,

Sexual offenses = .74, and Future plans = .78). The interrater reliability of the final risk judgment was moderate (ICC = .48). In 2 of 30 cases (6.7%), one rater judged “high risk” whereas another rater judged “low risk.” Two items—Sexual deviance and Relationship problems—demonstrated poor interrater reliability (ICCs = .38 and .29, respectively). For the item Relationship problems this was due to lack of variance. For the item Sexual deviance, we decided to further analyse the interrater reliability between the different raters. Because the first author (VdV) rated most of the cases, we computed ICCs between her scores and the scores of the other three raters. We found good reliability for VdV and CdR or DvB (ICC = .68, $N = 24$), but poor reliability for VdV and GM (ICC = .14, $N = 18$). In addition, the interrater reliability for GM and CdR or DvB was poor (ICC = $-.26$, $N = 24$). The fourth author (GM) was a master’s level student who only had had clinical experience with forensic cases during her internship, and this had an unfavorable impact on the overall level of reliability of item 1. Three items—Victim of child abuse, Employment problems, and Extreme minimization or denial of sex offenses—yielded moderate interrater reliability (ICCs = .49, and .48 and .42, respectively).

The overall interrater reliability of the Static-99 was excellent (ICC total score = .80). The risk category demonstrated good interrater reliability (ICC = .61). The ICC for item 6 could not be computed due to lack of variance, but the percentage of agreement was 98.9%.

Risk Judgments

The mean total scores on the SVR-20, Static-99, and PCL-R were 23.7 (SD 6.7), 6.0 (SD 1.7), and 21.9 (SD 7.2), respectively. There were no significant differences between the rapists and child molesters, except for the average number of other considerations in the SVR-20 (3.6 vs. 4.5, $p < .05$). Table 4 shows the final risk judgments of the SVR-20 and the risk categories of the Static-99. More than half of the group of sex offenders was judged to pose a “high risk” by both instruments. Furthermore, sexual recidivists obtained higher total scores on both instruments (SVR-20: 27.6 vs. 21.3; Static-99: 6.7 vs. 5.5, $p < .01$) and were more often judged to pose a high risk for sexual reoffending compared to sexual nonrecidivists (SVR-20: 91% vs. 30%; Static-99: 79% vs. 50%, $p < .01$). When a PCL-R total score of 30 was used as cut off score for the diagnosis of psychopathy, 21% of the sex offenders could be classified as psychopathic, and when 26 was used as cut off score 36%.

Table 4. Final Risk Judgment SVR-20/Risk Category Static-99

	Rapists $N = 95$	Child molesters $N = 27$	Total $N = 122$
SVR-20			
Low	17 (18%)	2 (7%)	19 (15%)
Moderate	27 (29%)	9 (33%)	36 (30%)
High	51 (54%)	16 (59%)	67 (55%)
Static-99			
Low	—	—	—
Medium low	5 (5%)	2 (7%)	7 (6%)
Medium high	36 (38%)	5 (19%)	41 (33%)
High	54 (57%)	20 (74%)	74 (61%)

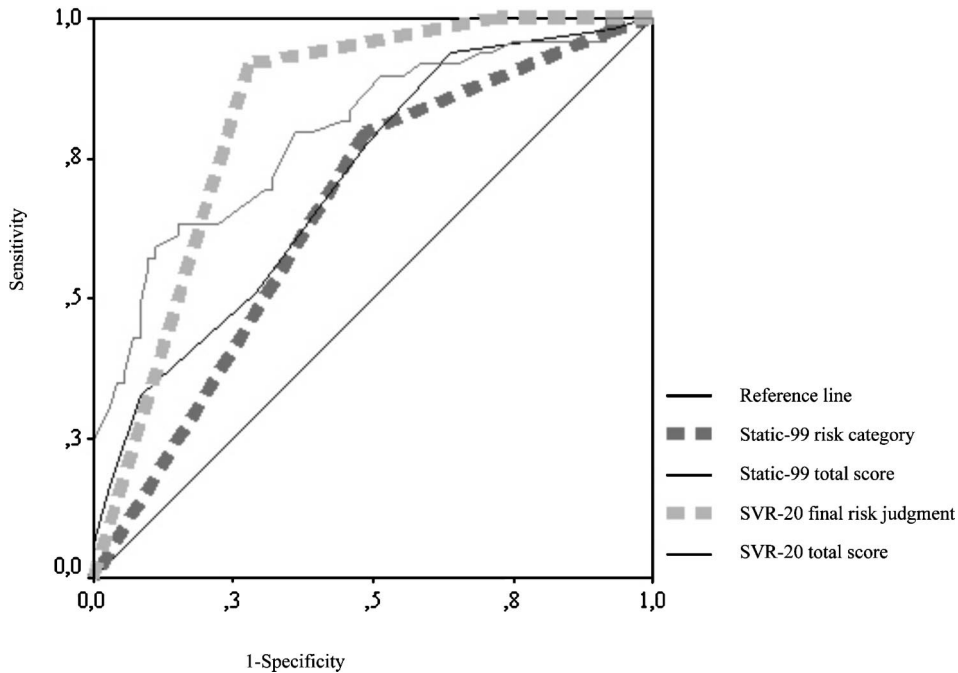


Fig. 1. ROC curves for SVR-20 and Static-99 (N = 121)

Predictive Validity

Figure 1 presents the ROC curve of the SVR-20 and Static-99 for sexual reoffending. The SVR-20 exhibited good predictive validity for sexual reoffending: all AUC values and Pearson *r*'s for the subscales, the total score, and the final risk judgment were significant (see Table 5). The total score, the final risk judgment, and the subscales Psychosocial adjustment and Future plans also predicted

Table 5. Predictive Validity of the SVR-20 and Static-99 (N = 121)

	Sexual recidivism N = 47			Violent offending N = 55			General offending N = 89		
	AUC	SE	r	AUC	SE	r	AUC	SE	r
SVR-20									
Psychosocial adjustment	.68***	.05	.30**	.67**	.05	.29**	.71***	.05	.33**
Sexual offenses	.79***	.05	.49**	.57	.05	.13	.60	.06	.16
Future plans	.76***	.04	.47**	.67**	.05	.32**	.73***	.05	.38**
Total score	.80***	.04	.50**	.66**	.05	.28**	.71***	.05	.33**
Final risk judgment	.83***	.03	.60**	.64*	.05	.26**	.69**	.05	.30**
Static-99									
Total score	.71***	.05	.38**	.54	.05	.11	.57	.06	.13
Risk category	.66**	.04	.30**	.50	.05	.03	.55	.06	.08

Note. AUC = Area under the curve. SE = Standard error. *r* = Pearson correlation coefficient. Violent offending = excluding sexual and homicide offenses.

* *p* < .05. ** *p* < .01. *** *p* < .001 (two-tailed).

violent (excluding sexual and homicide offenses) and general offending. The predictive validity of the Static-99 for sexual reoffending was moderate. The Static-99 was not predictive for violent (excluding sexual and homicide offenses) and general offending. However, when we included sexual offenses in the definition of violence, like Hanson and Thornton (2000) and Sjöstedt and Långström (2001) did in their studies, we found a significant predictive validity of the Static-99 total score for violent offenses ($AUC = .62, p < .05$). When we used AccuROC to compare the AUC values of the SVR-20 and Static-99, we found a marginally significant difference between the AUC values of the SVR-20 total score and Static-99 total score ($\chi^2 = 2.8, df = 1, p = .09$). The difference in AUC values of the SVR-20 final risk judgment and Static-99 risk category, however, was significant ($\chi^2 = 15.0, df = 1, p < .001$).

Other Considerations

Other considerations that were frequently coded on the SVR-20 are lack of coping skills, suggestibility, impulsivity, failure of prior treatment(s), social isolation, lack of social and emotional support, financial problems, and preoccupation with nondeviant sex (hypersexuality).

DISCUSSION

The present study is the first one in The Netherlands to assess the predictive validity of risk assessment instruments for sexual violence. The base rates for sexual recidivism found in this study are comparable to those found in other studies with long follow-up periods (e.g., Prentky et al., 1997). Furthermore, the present study yielded good interrater reliability for the SVR-20 total and subscale scores and the Static-99 items and total score. However, for the final risk judgment and some individual SVR-20 items we found only a moderate or—for two items—poor interrater reliability. An explanation for this was lack of information in the files to code specific items, for example, item 17 (Extreme minimization or denial of sex offenses). In addition, the description of some items is rather broad and therefore open to multiple interpretations. In particular, coding item 1 (Sexual deviance) for rapists raised a lot of discussion because clear and objective criteria for sexual deviance are lacking. This is especially true for forensic assessment in The Netherlands where phallometric methods to assess sexual deviance are not generally used.

Regarding the predictive accuracy, we found a difference between the actuarial instrument and the SPJ guideline. Although the difference in AUC values of the SVR-20 total score and Static-99 total score was only marginally significant, the SVR-20 final risk judgment was significantly more accurate in predicting sexual recidivism than the Static-99 risk category. Although both instruments were specifically designed to predict sexual violence, it turned out that the SVR-20 also had significant predictive accuracy for nonsexual violent and general offenses. The SVR-20 subscale Sexual offenses was specifically predictive for sexual reoffending, whereas the subscales Psychosocial adjustment, and Future plans also predicted nonsexual

violent and general offenses. Interestingly, the AUC value and Pearson correlation for the SVR-20 final risk judgment were both the highest observed. Thus, using the SVR-20 as a SPJ method seems to have superior accuracy than using it as an actuarial tool (i.e. summing the individual item scores). This finding is in line with Dempster (1998) who found incremental predictive validity for the SVR-20 structured final risk judgment relative to the actuarial SVR-20 (i.e., the SVR-20 total score). The same was found for the Spousal Assault Risk Assessment guide (SARA; Kropp, Hart, Webster, & Eaves, 1999), a SPJ guideline for the assessment of domestic violence (Kropp & Hart, 2000), and the HCR-20. Douglas, Ogloff, and Hart (2003) demonstrated in a sample of 100 forensic psychiatric that the SPJ structured final risk judgments added incremental validity to the HCR-20 used in a numerical sense.

The Static-99 exhibited a moderate predictive accuracy for sexual reoffending, and no significant predictive validity for nonsexual violent or general offenses. The predictive validity for sexual reoffending resembles that documented by Hanson and Thornton (2000) and Sjöstedt and Långström (2001). Contrary to their findings, we found no predictive validity for nonsexual violent and general offenses. However, these authors adopted a different definition of violent offenses, which included sexual offenses. We decided to exclude sexual offenses because we wanted to make a clear distinction between sexual and nonsexual violent offenses. When we included sexual offenses in our violent recidivism definition, we also found a significant predictive validity of the Static-99 for violent offending. Furthermore, the Static-99 risk categories did not differentiate in our sample: 94% of the sex offenders were classified as medium high or high by this instrument. Perhaps the Static-99 will show better differentiation and predictive accuracy in other populations, such as sex offenders in outpatient settings. We suggest that—especially for settings where time, staff, and information about patients is limited—the Static-99 could serve as an instrument for a first, global screening of sexual recidivism risk to decide whether more elaborate risk assessment with the SVR-20 is desirable. There are a number of limitations to the present study. The first limitation has to do with a possible rater effect, which may subsequently influence the generalizability of the findings. Only four raters participated in this study, of which rater 1 (VdV) coded the majority of the cases. We believe this is a greater limitation for the coding of the SVR-20. The Static-99 is rather straightforward to code and the final risk category is the result of a fixed algorithm, thus, we do not believe that the rater effect can be strong for this instrument. This hypothesis is confirmed by the excellent interrater reliability of the Static-99 and the more moderate reliability of the SVR-20 final risk judgment. Another issue concerning the raters is that some of them may have held (strong) views regarding the superiority of the SPJ method to the actuarial method, which may have impacted the results. However, all raters were blind to the recidivism data when they coded the patient file data and it is unlikely that a positive expectancy regarding the SVR-20 alone, could account for the large difference in predictive validity between the two instruments. In future research, numerous different raters who are blind to the outcome and to the hypothesis of the study are needed. Second, the sex offenders formed a select group, because they had severe psychological problems and had committed serious offenses for which involuntary treatment was considered necessary.

Although this group of sex offenders is representative for Dutch sex offenders with a tbs-order (van Emmerik & Brouwers, 2001), they are probably not for sex offenders in general. Both the SVR-20 and Static-99 show that the majority of the sex offenders in our study are a high-risk group. Third, 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 research has revealed that many sexual offenses go undetected and not all sex offenders are apprehended and arrested (Groth, Longo, & McFadin, 1982; Weinrott & Saylor, 1991). Moreover, the Judicial Documentation data cannot be considered as completely reliable for long term follow-up studies like the present study, because—as stated by Dutch Criminal law—offenses that occurred 20 years or longer ago have to be removed from the register (Dutch Criminal Code, Act of Judicial Documentation, section 7). Lastly, there are limitations that relate to the retrospective design of this study. The quality of the files varied substantially. Since the 1990s, treatment progress was documented much more carefully than in the seventies and eighties, which may have influenced coding. This may have especially affected coding the SVR-20 items, because these items comprise quite complex constructs such as sexual deviance and escalation in sexual offending. In contrast, the Static-99 items are relatively easy to code from file information. Furthermore, some of the sex offenders were subjected to outdated treatment methods that do not correspond to current best practice.

Future research will have to focus on groups of sex offenders across different settings and contexts, for instance, in the prison system and outpatient settings. Prospective research is recommended, although a number of problems might be encountered. The most important problem is that prospective predictive research will be hampered by the clinical goals of risk assessment, i.e., risk management and prevention. 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” (Hart, 1998, p. 365). Thus, when clinicians perform SVR-20 or HCR-20 risk assessments it is very likely that the outcome influences decisions concerning probationary leave or termination of (mandatory) treatment and high-risk patients will not be released from the hospital. Therefore, retrospective studies such as the present study are particularly suitable to assess psychometric properties, most notably their predictive validity, of risk assessment instruments.

Risk assessment instruments should be regarded as “work in progress” and further improvement of these instruments is desirable (Webster et al., 1997b). More specifically, attention needs to be paid to the development and refinement of dynamic risk factors and protective factors, as well as theoretical models that explain the relationship between risk factors and actual offending. Our study has directed attention to a number of other risk factors, some of which might be valuable additions to the SVR-20, for instance, hypersexuality and social isolation. Finally, we should bear in mind that one of the most important goals of structured risk assessment is to gain insight into strategies to diminish risk for sexual and violent behavior.

ACKNOWLEDGMENT

The authors thank Ine Kusters, M.Sc. and Quinta Appeldoorn for their assistance in retrieval of archival hospital data.

REFERENCES

- Belfrage, H., Fransson, G., & Strand, S. (2000). Prediction of violence using the HCR-20: A prospective study in two maximum-security correctional institutions. *The 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, British Columbia: Institute Against Family Violence.
- Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines and training. *American Psychologist, 51*, 945–956.
- Cooke, D., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment, 13*, 171–188.
- 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.
- Dempster, R. J. (1998). *Prediction of sexually violent recidivism: A comparison of risk assessment instruments*. Unpublished master's thesis, Simon Fraser University, Vancouver, British Columbia, Canada.
- Dempster, R. J., & Hart, S. D. (2002). The relative utility of fixed and variable risk factors in discriminating sexual recidivists and nonrecidivists. *Sexual Abuse: A Journal of Research and Treatment, 14*, 121–138.
- Doren, D. M. (1998). Recidivism base rates, predictions of sex offender recidivism, and the “sexual predator” commitment laws. *Behavioral Sciences and the Law, 16*, 97–114.
- Douglas, K. S. (2001). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*. Available from <http://www.sfu.ca/psychology/groups/faculty/hart/violink.htm>.
- Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment. *Criminal Justice and Behavior, 29*, 617–658.
- 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., 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.
- Douglas, K. S., & Webster, C. D. (1999). Predicting violence in mentally and personality disordered individuals. In R. Roesch, S. D. Hart, & J. R. P. Ogloff (Eds.), *Psychology and law: The state of discipline* (pp. 175–239). New York: Kluwer Academic.
- Grann, M., Långström, N., Tengström, A., & Stålenheim, E. G. (1998). The reliability of file-based retrospective ratings of psychopathy with the PCL-R. *Journal of Personality Assessment, 70*, 416–426.
- Greenberg, D. M. (1998). Sexual recidivism in sex offenders. *Canadian Journal of Psychiatry, 43*, 459–465.
- Groth, A. N., Longo, R. E., & McFadin, J. B. (1982). Undetected recidivism among rapists and child molesters. *Crime and Delinquency, 28*, 450–458.
- Grubin, D. (1998). Sex offending against children: Understanding the risk. In *Police Research Series Paper 99*. London: Home Office.
- Grubin, D., & Wingate, S. (1996). Sexual offence recidivism: Prediction versus understanding. *Criminal Behaviour and Mental Health, 6*, 349–359.
- Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York: Wiley.
- Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism* (User Report 97-04). Ottawa: Department of the Solicitor General of Canada.
- Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348–362.
- Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior, 27*, 6–35.

- Hanson, R. K., Steffy, R. A., & Gauthier, R. (1993). Long-term recidivism of child molesters. *Journal of Consulting and Clinical Psychology, 61*, 646–652.
- Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders* (User Report No.1999-02). Ottawa: Department of the Solicitor General of Canada.
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessment for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*, 119–136.
- 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.
- Harris, G. T., & Rice, M. E. (1997). Risk appraisal and management of violent behavior. *Psychiatric Services, 48*, 1166–1176.
- 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. (Ed.). (2002). Swedish studies on psychology, crime and law [Special issue]. *Psychology, Crime, and Law, 8*.
- 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, The Netherlands: Kluwer.
- Hildebrand, M., de Ruiters, 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.
- Hildebrand, M., de Ruiters, 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 Ruiters, C., & de Vogel, V. (2004). Psychopathy and sexual deviance in treated rapists: Association with (sexual) recidivism. *Sexual Abuse: A Journal of Research and Treatment, 16*, 1–24.
- Hildebrand, M., de Ruiters, 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.
- 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.
- 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.
- Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law, 7*, 409–443.
- 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. (1995). Correlates of accuracy in the assessment of psychiatric inpatients' risk of violence. *American Journal of Psychiatry, 152*, 901–906.
- Monahan, J. (1981). *The clinical prediction of violent behavior*. Rockville, MD: National Institute of Mental Health.
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783–792.
- Prentky, R. A., Lee, A. F. S., Knight, R. A., & Cerce, D. (1997). Recidivism rates among child molesters and rapists: A methodological analysis. *Law and Human Behavior, 21*, 635–658.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Quinsey, V. L., Lalumière, M. L., Rice, M. E., & Harris, G. T. (1995). Predicting sexual offenses. In J. C. Campbell (Ed.), *Assessing dangerousness: Violence by sex offenders, batterers, and child abusers* (pp. 114–137). Thousand Oaks, CA: Sage.
- Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737–748.
- Rice, M. E., & Harris, G. T. (1997). Cross-validation and extension of the Violence Risk Appraisal Guide for child molesters and rapists. *Law and Human Behavior, 21*, 231–241.
- 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.

- Schmidt, P., & Wytte, A. D. (1988). *Predicting recidivism using survival models*. New York: Springer.
- Sjöstedt, G., & Långström, N. (2001). Actuarial assessment of sex offender recidivism risk: A cross-validation of the RRASOR and the Static-99 in Sweden. *Law and Human Behavior, 25*, 629–645.
- Sjöstedt, G., & Långström, N. (2002). Assessment of risk for criminal recidivism among rapists: A comparison of four different measures. *Psychology, Crime, and Law, 8*, 25–40.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Allyn & Bacon.
- van Beek, D. J. (1999). *De delictscenarioprocedure bij seksueel agressieve delinquenten* [The offense chain procedure with sexually aggressive offenders]. Arnhem, The Netherlands: Gouda Quint.
- van Beek, D. J., de Doncker, D., & de Ruiter, C. (2001). *Static-99. Inschatten van het risico van seksueel gewelddadige recidive bij volwassen seksuele delinquenten* [Static-99. Assessment of the risk of sexually violent recidivism in adult sex offenders]. Utrecht, The Netherlands Forum Educatief.
- 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 Netherlands Dienst Justitiële Inrichtingen.
- 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, The Netherlands: Swets Test Publishers.
- Vida, S. (1997). *AccuROC Nonparametric Receiver Operating Characteristic Analysis* (Version 3.1) [Computer software]. Montreal, Quebec, Canada: Accumetric Corporation.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997a). Assessing risk of violence to others. In C. D. Webster & A. Jackson (Eds.), *Impulsivity. Theory, assessment and treatment* (pp. 251–272). New York: The Guilford Press.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997b). *HCR-20. Assessing the risk of violence. Version 2*. Burnaby, British Columbia, Canada: Simon Fraser University and Forensic Psychiatric Services Commission of British Columbia.
- Weinrott, M. R., & Saylor, M. (1991). Self report of crimes committed by sex offenders. *Journal of Interpersonal Violence, 5*, 283–300.