

SAVRY RISK ASSESSMENT IN VIOLENT DUTCH ADOLESCENTS

Relation to Sentencing and Recidivism

HENNY P. B. LODEWIJKS

Rentray Juvenile Correctional and Treatment Facility

THEO A. H. DORELEIJERS

VU University Medical Center

CORINE DE RUITER

Maastricht University

This study examines the predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) by examining relationships between SAVRY scores and violent reoffending during a 3-year period after sentencing. Two types of sentences were studied: a mandatory treatment order ($N = 77$) and a juvenile prison sentence ($N = 40$). The predictive validity of the SAVRY was significant for the two types of sentences. The predictive validity of the unstructured clinical judgment proved to be not significant. Support was found for the hypothesis that the juvenile court's sentence (treatment versus detention) might have been influenced by the unstructured clinical risk assessment of the mental health experts, even though this assessment is a poor predictor of violent reoffending.

Keywords: *sentencing; youth violence; risk assessment; SAVRY; predictive validity*

Assessment of violence risk among adolescents is a critical and essential part of juvenile justice interventions. In the past decade, considerable advances have been made in violence risk assessment research and clinical practice among juvenile offenders (Borum & Verhaagen, 2006; Lodewijks, Doreleijers, de Ruiters, & Borum, in press; Schmidt, Hoge, & Gomes, 2005). Research identifying factors associated with increased risk of violent offending in juveniles has been applied to develop instruments for standardized risk assessment. These structured assessments have been found to provide a more valid and consistent assessment of criminogenic risk and need than unstructured assessments (Hoge, 2002). These risk assessment tools for juveniles were predominantly developed in North America (Bartel, Borum, & Forth, 2000; Hoge & Andrews, 2002). Very few instruments are available that are specifically designed for a comprehensive assessment of risk for future violence in youth. A widely used instrument for adolescents, the Youth Level of Service/Case Management Inventory (YLS/CMI; Hoge & Andrews, 2002), for instance, has general criminal and delinquent offending and not specifically violence as risk outcome.

AUTHORS' NOTE: *The Dutch Ministry of Justice and the Rentray Foundation financially supported this study. The authors wish to thank the Forensic Psychiatric Service for providing access to their files. Special thanks go to Henny de Wit-Grouls, who participated as a research assistant in the study. Address correspondence to Henny Lodewijks, P.O. Box 94, 7200 AB Zutphen, The Netherlands; e-mail: hlodewijks@rentray.nl.*

CRIMINAL JUSTICE AND BEHAVIOR, Vol. XX No. X, Month XXXX xx-xx

DOI: 10.1177/0093854808316146

© International Association for Correctional and Forensic Psychology

In the past two decades, the objective of risk assessment changed from a pure risk-prediction approach to a more clinically relevant risk assessment and risk management model (Monahan, 2003; Webster, Hucker, & Bloom, 2002). In line with this change, especially in juveniles, risk assessment should emphasize changeable, dynamic risk factors (Borum & Verhaagen, 2006; Steinberg & Schwartz, 2000). Recently, limits of forensic mental health assessment for accurate prediction of risk of violence have been reported for child sexual abuse (Herman, 2005) and for violence related to posttraumatic stress disorder (Koch, O'Neill, & Douglas, 2005).

RISK ASSESSMENT INSTRUMENTS

Risk assessment instruments can be divided into actuarial and structured professional judgment (SPJ) instruments. In several studies, the SPJ instruments have outperformed the actuarial instruments of risk assessment, both in research with adults (Dempster, 1998; de Vogel, de Rooter, Beek, & Mead, 2004) and adolescents (Bartel et al., 2000). In both the actuarial and SPJ models, the evaluator systematically assesses a set of predetermined, empirically derived risk factors. The main difference between the actuarial and the SPJ approach is in how the final risk judgment is made. For actuarial instruments, the decision is made according to a fixed algorithm. For SPJ instruments, the professional makes the final risk judgment based on the structured assessment of all risk factors (Otto, 2000). In the SPJ method, the professional not only rates and sums the items but also uses personal expertise and knowledge to weigh, combine, and interpret the risk factors to arrive at a final risk judgment.

The SAVRY is an SPJ instrument. The structure of the SAVRY is modeled on existing risk assessment guidelines for adults such as the Historical, Clinical Risk Management–20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997), but its item content focuses on risk factors relevant to adolescents. The SAVRY guideline consists of 24 risk items, divided into three domains (historical, social/contextual, and individual) and one protective domain with six items (see Table 1). The risk items have a three-level coding structure (low, moderate, high) and the protective items have a two-level structure (absent or present). Specific coding guidelines for each item and each level are provided in the manual.

The SAVRY manual explicitly advises against the use of numerical indices and cutoff points in clinical decision making. Only for research purposes is the SAVRY risk total score used. The SAVRY risk total score is derived by numerically transforming and summing codes of low, moderate, and high on the 24 risk items to 0, 1, and 2, respectively. In clinical applications, the summary risk rating is used. The SAVRY summary risk rating is the final professional judgment, based on an overall interpretation of the 24 risk items and the 6 protective items. This summary rating is not linked to a particular score or range of scores.

Psychometric support for the SAVRY is presented in the manual (Borum, Bartel, & Forth, 2002) and on the Web site (www.fmhi.usf.edu/mhlp/savry/statement.htm). In the initial validation study (Bartel et al., 2000), SAVRY risk total scores were significantly related to measures of institutional aggressive behavior ($r = .40$) and conduct disorder symptoms ($r = .52$). Protective factors were negatively related to these outcomes. In other studies, significant correlations have also been found between SAVRY risk total scores and measures of violence among young male offenders in Canada (Catchpole & Gretton, 2003; Gretton & Abramowitz, 2002) and high-risk Native American youth (Fitch, 2002). SAVRY summary risk ratings have also been found to correlate significantly with outcome measures of community violence (Gretton & Abramowitz, 2002; McEachran, 2001). Using receiver

TABLE 1: Items of the Structured Assessment of Violence Risk in Youth (SAVRY)

<i>Historical Items</i>		<i>Social/Contextual Items</i>	
1.	History of violence	11.	Peer delinquency
2.	History of nonviolent offending	12.	Peer rejection
3.	Early initiation of violence	13.	Stress and poor coping
4.	Past supervision/intervention failures	14.	Poor parental management
5.	History of self-harm or suicide attempts	15.	Lack of personal/social support
6.	Exposure to violence in the home	16.	Community disorganization
7.	Childhood history of maltreatment		
8.	Parental/caregiver criminality		
9.	Early caregiver disruption		
10.	Poor school achievement		
<i>Individual Items</i>		<i>Protective Items</i>	
17.	Negative attitudes	P1.	Prosocial involvement
18.	Risk taking/impulsivity	P2.	Strong social support
19.	Substance use difficulties	P3.	Strong attachments and bonds
20.	Anger management problems	P4.	Positive attitude toward intervention and authority
21.	Low empathy/remorse	P5.	Strong commitment to school or work
22.	Attention deficit/hyperactivity difficulties	P6.	Resilient personality
23.	Poor compliance		
24.	Low interest/commitment to school or work		

Note. All historical, social/contextual, and individual items are coded on a three-level scale: low, moderate, or high. All protective items are coded on a two-level scale: absent or present. The summary risk rating is to be made as a professional judgment based on all risk and protective items in the case (from Borum et al., 2002).

operating characteristic (ROC) analysis, areas under the curve (AUCs) for the SAVRY risk total score average between .74 and .80 across these studies. Interestingly, the assessors' overall risk judgment (summary risk rating) consistently performs as well as, and often better than, the actuarial combination of the scores. For example, using ROC analysis, McEachran (2001) found an AUC for the SAVRY risk total score of .70, but the AUC for the SAVRY summary risk rating was .89. Lodewijks, Doreleijers, de Ruiters, and Borum (in press) also found a difference, although not significant, in favor of the summary risk rating.

The SAVRY is a North-American-based tool and its appropriateness in non-North American samples needs to be demonstrated. Cross-cultural differences could affect the validity of instruments. For instance, the United States has the highest youth homicide rate among the 26 wealthiest nations (American Academy of Pediatrics, 1999). Compared to European countries, the United States also has more gun-related violence (Sheppard, Rowe, Grant, & Jacobs, 2003) and more violence related to gang membership (Howell & Egley, 2005).

DUTCH JUVENILE CRIMINAL LAW

Under Dutch criminal law, a juvenile violent offender can be sentenced to a mandatory treatment measure or to detention or given an alternative sentence that can take the form of community service (van der Laan, 2006). The mandatory treatment measure and detention are carried out in a juvenile justice facility, although on separate wards. Unique to The Netherlands is the possibility that a young, seriously delinquent offender in need of intensive treatment can be sentenced to compulsory treatment in a juvenile justice institution. In cases of serious violence, it is common for the juvenile judge to request a forensic mental

health assessment to help decide whether detention or a mandatory treatment measure is appropriate. This mental health assessment should not only address mental health problems but also risk for future violence. Until recently, the assessment of risk in juvenile offenders in The Netherlands was based on unstructured clinical judgment (Duits, 2006).

It is quite common in forensic assessment for risk of future violence to be stated in terms of what might happen if no intervention would take place. For example, when risk of future violence is judged as high by the professional who made the assessment, the judge will more likely impose a mandatory treatment order because of the implicit assumption that treatment can reduce recidivism risk. When risk of future violence is evaluated as low, the judge will more likely impose a detention order, mostly comparable to the remand period. Judges usually follow the advice made in the mental health assessment (Duits, 2006).

The type of sentence has far-reaching consequences for the juvenile in terms of duration of incarceration and whether or not treatment will be provided. A detention sentence can last up to 1 year between the ages of 12 and 16 and up to 2 years between 16 and 18. The mandatory treatment order is generally considered a more intensive intervention than mere detention because of a longer stay in a treatment facility. A mandatory treatment order can be imposed between the ages of 12 and 18, and it can vary in duration from 2 to 6 years depending on the seriousness of the offense and on whether a mental disorder was present at the time of the offense.

Wartna, Kalidien, Tollenaar, and Essers (2006) found that general recidivism in adolescents in The Netherlands within a follow-up period of 4 years ranged from 62% to 74%, depending on whether the adolescent was sent to a mandatory treatment ward or a detention ward. When the index offense included only serious offenses, the general recidivism rate within 4 years after discharge was 70% and the violent recidivism rate was 25% (van der Heiden-Attema & Wartna, 2000).

HYPOTHESES AND RESEARCH QUESTIONS

Several authors (Borum, 2003; Hogan & Campbell, 2005) have argued for the performance of systematic and developmentally informed risk assessments in combination with the selective assignment of interventions to young offenders. One such risk assessment instrument is the SAVRY. Although past research findings with this instrument have been promising, further research is needed to study the predictive validity of the SAVRY with regard to court-ordered assessments, especially for the Dutch language version of the SAVRY (Lodewijks et al., 2003).

In the present study, the strength of the association between the SAVRY and violent outcome is examined in a sample of Dutch adolescents who were brought to justice because of a violent offense. At the time of the unstructured clinical judgment, between 1998 and 2002, the SAVRY was unknown in The Netherlands. To enable a comparison between structured and unstructured clinical judgment, the ratings of the SAVRY were based on the same file information as was available at the time of the unstructured clinical judgment. On the basis of prior research, we hypothesized the following:

Hypothesis 1: The SAVRY summary risk rating and the SAVRY risk total score would have good interrater reliability and good predictive validity for violent reoffending.

Hypothesis 2: The SAVRY summary risk rating would significantly outperform the SAVRY risk total score in terms of predictive value.

Hypothesis 3: The SAVRY summary risk rating and the SAVRY risk total score would significantly outperform unstructured clinical judgment.

Furthermore, we had the following research questions about the sentence:

Research Question 1: Would there be a difference in SAVRY outcomes related to the type of sentence?

Research Question 2: Would there be a difference in recidivism rate between the two sentences?

Research Question 3: Would the juvenile judge's sentencing decision be more influenced by the unstructured clinical judgment expressed in the mental health evaluation than by the nature of the index offense?

METHOD

PARTICIPANTS

The sample consisted of 117 juvenile offenders with a mean age of 15.3 years at the time of assessment ($SD = 1.3$, range = 12 to 18 years). These youths were taken from a total sample of 130 violent juvenile offenders prosecuted in Amsterdam, The Netherlands between October 1998 and May 2002. They were evaluated by a psychologist and, in most of the cases, also by a psychiatrist at the request of the juvenile court prior to sentencing. Thirteen adolescents were excluded from the sample either because the SAVRY could not be rated because of incomplete file information ($n = 5$) or because they had not had any unsupervised leaves during the follow-up period of 3 years, so their time at risk was zero ($n = 8$).

Table 2 presents demographic characteristics of the total sample, all having had committed violent offenses. The mandatory treatment lasted 2 years (14%, $n = 10$), 4 years (25%, $n = 17$), or 6 years (61%, $n = 17$). Depending on the recidivism risk, as judged by the treatment institution, the judge decides every 2 years whether the mandatory treatment needs to be extended (of course, this applies only to the 4-year and 6-year orders). During the follow-up period, the mandatory treatment was terminated in 11 cases because of the end of the 2-year period and was extended for half a year in 10 cases. At the end of the 3-year period, 56 youths still were in a juvenile justice institution: 12 in a secure and 44 in a semi-secure institution. The actual average stay in an institution for mandatory treatment during the 3-year follow-up period, either secure or semi-secure, was 1,031 days ($SD = 129$, range 593-1,095). Detention lasted an average of 76 days ($SD = 74$, range 11-358). All time periods were reckoned from the date of the forensic mental health reports. The difference between mandatory treatment and detention was significant, $\chi^2(1, 54) = 117, p < .001$.

PROCEDURE

Four raters, all master's-level psychologists, were trained in coding the SAVRY during a 2-day workshop given by the first author, a senior clinical psychologist. This workshop reviewed the relevant empirical literature and provided practice cases for coding the SAVRY using file information and videos of actual cases.

Interrater reliability was calculated based on 25 randomly selected cases rated independently by two raters. Each rater-pair combination rated an equal number of cases. Subsequently, the two raters discussed their ratings and agreed on a consensus rating on the individual items and on the final risk judgment. After the training and consensus meetings, two raters coded the other files. The consensus SAVRY ratings ($n = 50$) and the single rated SAVRY's ($n = 67$) were used for subsequent analyses of the predictive validity.

TABLE 2: Sample Characteristics (N = 117)

<i>Characteristics</i>	n	%
Gender		
Male	111	95
Female	6	5
Ethnic origin		
Caribbean (Surinamese, Antillean)	48	41
Mediterranean (Moroccan, Turkish)	27	23
Caucasian (Dutch and other European countries)	33	28
Other descent	9	8
Index offense		
Violent property offense	64	55
Manslaughter and (aggravated) assault	40	34
Sexual offense	12	10
Arson with danger of persons	1	1
Sentence		
Mandatory treatment	77	66
Detention	40	34

Both the unstructured and structured assessment made use of the same files. The raters who used the SAVRY were unaware of the final advice to the judge based on the unstructured risk assessment. As mentioned previously, the SAVRY was not available at the time of the unstructured risk assessment and, therefore, was used retrospectively. Unstructured clinical judgment was based on a review of the concluding comments of the forensic mental health assessment reports by an experienced forensic psychologist (for a similar procedure: see de Vogel, de Ruiter, Hildebrand, Bos, & van de Ven, 2004). This forensic psychologist was uninformed of the type of sentence and actual recidivism of the offender. Raters of the SAVRY were blind to the unstructured clinical judgment, type of sentence, and recidivism data.

VIOLENCE OUTCOME

Data on violent recidivism of the subjects were retrieved from the Judicial Documentation Register of the Ministry of Justice. This outcome measure is by far the most frequently used in recidivism studies in The Netherlands (de Vogel & de Ruiter, 2006; Philipse, 2005). In accordance with Dutch criminal law, recidivism was defined as a new conviction by the court for an offense. For the identification of violent offenses, the SAVRY violence definition was applied: “an act of battery or physical violence that is sufficiently severe to cause injury to another person or persons (i.e., cuts, bruises, broken bones, death, etc.), regardless of whether injury actually occurs; any act of sexual assault; or a threat made with a weapon in hand” (Borum et al., 2002, p. 29).

Recidivism was measured during the 3 years after the date of the forensic mental health assessment. Offenses during the stay in the institution were included when they had occurred during unsupervised leaves or during absconding. Institutional violent offenses were not considered because the juvenile treatment institutions differ substantially in their policies and practices regarding the reporting of institutional offenses to the justice authorities, and thus figures are highly divergent. Time at risk for the mandatory treated youth was calculated by adding the number of days of unsupervised leave, escape from the institution, not returning from leave, probationary leave, and the period after discharge (until, for some

cases, the first sentence of renewed placement in an institution). Time at risk for the detention youth was the number of days after release and, for some cases, until the first sentence of renewed placement in an institution.

STATISTICAL ANALYSES

To test Hypothesis 1, the interrater reliability was assessed by means of the intraclass correlation coefficient (ICC), using the two-way, random effects variance model and consistency type (McGraw & Wong, 1996). We used the following critical values for single measure ICC's: $ICC \geq .75 = \textit{excellent}$; ICC from $.60$ to $.75 = \textit{good}$; ICC from $.40$ to $.60 = \textit{moderate}$; $ICC < .40 = \textit{poor}$ (Fleiss, 1986). To test Hypothesis 1 on validity, the predictive validity was assessed by means of ROC analysis (Mossman, 1994; Rice & Harris, 1995). We chose this statistical method because it is less reliant than other statistical analyses (like correlation coefficients) on base rates for recidivism and the particular cut-off score chosen to classify cases. Also, normality need not be assumed (Rice & Harris, 1995). ROC analyses result in a plot of the true positive rate (sensitivity) against the false positive rate ($1 - \textit{specificity}$) for every possible cut-off score of the instrument. The resulting AUC can be interpreted as the probability that a randomly selected reoffender would score higher on the instrument than a randomly selected nonreoffender. An AUC of $.50$ represents chance prediction, and an AUC of 1.0 perfect prediction. In general, AUC values of $.70$ and greater are considered *moderate* and those greater than $.75$, *good* (Douglas, Guy, & Weir, 2005). To test Hypotheses 2 and 3, to compare the obtained AUC values, we used AccuROC version 2.5 (Vida, 1997). Furthermore, Cox regression analysis (Cox, 1972) was conducted to evaluate whether the risk total score and summary risk rating add incremental value to unstructured clinical judgment as a predictor of violent recidivism. Spearman ρ correlations were calculated to facilitate comparison with other studies that also provided correlations.

To answer Research Question 1, Student's t test was used to examine the differences between the two sentence types and SAVRY outcomes. To answer Research Question 2, survival analysis, referred to as the Kaplan–Meier method, was used to calculate recidivism rates and the average time prior to reoffending. Survival analysis analyzes the time to event, is able to account for censoring, and calculates the probability of recidivating for each time period in which the offender has not yet reoffended (Hosmer & Lemeshow, 1999). Once an offender recidivates, he is removed from the analysis for the subsequent time period. The log rank statistic was used to test the difference between the survival curves of the two groups (mandatory treatment and detention). To answer Research Question 3, to calculate the influences by the unstructured clinical judgment and by index offense on the type of sentencing, we used chi-square analysis. All statistical analyses were conducted using SPSS version 15.0.

RESULTS

INTERRATER RELIABILITY (HYPOTHESIS 1)

The single measure ICC demonstrated excellent interrater reliability for the SAVRY summary risk rating ($.82$) and the risk total score ($.80$). ICC's for the four domains separately ranged from *good* to *excellent* (Historical = $.74$, Social/Contextual = $.62$, Individual = $.82$, and Protective = $.86$).

No systematic bias was found in total risk scores or final risk judgments or in the unstructured risk assessment with regard to gender, age or ethnic origin.

PREDICTIVE VALIDITY (HYPOTHESES 1, 2, AND 3)

Tables 3, 4, and 5 show the AUC values of the SAVRY domains, risk total score and summary risk rating and of the unstructured clinical judgment. For all participants together and for the two sentence types, the risk total score and summary risk rating were significantly greater than .50. However, the historical domain never yielded a significant prediction. All other domains were significantly predictive of future violent offending. The unstructured clinical judgment never outperformed chance prediction. These results are in line with our hypotheses. We added the confidence intervals to facilitate comparison with other studies that also provided confidence intervals. As far as we know, the intervals found in our study are quite common in risk assessment research.

When we compared the AUC values for violent recidivism with AccuROC, we found a significant difference between the unstructured and structured risk judgments, z statistic = 2.8, $p = .005$, two-tailed. The difference between the AUC's of the total risk score and the summary risk rating was not significant, z statistic = .7, $p = .47$, two-tailed.

To facilitate comparison with other studies in this area, we calculated Spearman ρ correlations and found significant relations between the summary risk rating and violent recidivism for both groups (mandatory treatment: $\rho = .27$, $p < .05$; detention: $\rho = .44$, $p < .01$).

SAVRY OUTCOMES (RESEARCH QUESTION 1)

The highest rates of missing values occurred for Item 15 (6%; "lack of personal/social support"), Item 16 (6%; "community disorganization"), and Item 23 (5%; "poor compliance"). The three most frequently coded other considerations were: mentally retarded (10) and victim of sexual abuse outside the family (8) in the historical domain, and problems with acculturation (3) in the social/contextual domain.

Table 6 presents mean scores and standard deviations for the SAVRY domains, risk total scores and summary risk rating for both groups. As can be seen from this table, the mean SAVRY domains and the risk total score differ significantly between the two groups. Youth sentenced to mandatory treatment were judged as being at significantly higher risk. Mandatory-treatment youth scored significantly higher on all risk domains and significantly lower on the protective domain than did detention youth.

FOLLOW-UP AND REOFFENDING (RESEARCH QUESTION 2)

The average time at risk during the follow-up period for the mandatory-treatment-sentenced youth was 80 days ($SD = 146$, range 10-649). The average time at risk for the detention-sentenced youth was 1,031 days ($SD = 195$, range 411-1,095).

A total of 11 (14%) of the mandatory-treatment-sentenced youth and 12 (30%) of the detention-sentenced youth were reconvicted of violent offenses. There were significant differences in the failure rates, computed with survival analysis, between the two groups on violent recidivism, log rank (1, 117) = 5.1, $p < .05$. All reoffenses could be categorized as serious according to Loeber, Farrington, and Waschbusch (1998, p. 15): attempted homicide (2), aggravated assault (11), robbery (6), manslaughter (3), and rape (1). Although not part

TABLE 3: Predictive Validity of the SAVRY for Violent Reoffending Regardless of Type of Sentence (N = 117; follow-up 3 years)

	<i>AUC</i>	<i>SE</i>	<i>CI 95%</i>
Historical domain	.49	.07	.36-.62
Social/contextual domain	.67*	.06	.55-.79
Individual domain	.68**	.07	.55-.81
Risk total	.65*	.06	.53-.81
Protective domain	.28***	.06	.17-.40
Summary risk rating	.71**	.06	.60-.82
Unstructured clinical judgment	.45	.07	.34-.60

Note. SAVRY = Structured Assessment of Violence Risk in Youth; AUC = area under the curve; SE = standard error; CI = confidence interval.

* $p < .05$. ** $p < .01$. *** $p < .001$, two-tailed.

TABLE 4: Predictive Validity of the SAVRY for Violent Reoffending for the Mandatory Treatment Group (N = 77, follow-up 3 years)

	<i>AUC</i>	<i>SE</i>	<i>CI 95%</i>
Historical domain	.51	.08	.35-.67
Social/contextual domain	.74**	.07	.60-.87
Individual domain	.72*	.08	.55-.89
Risk total	.70*	.07	.57-.84
Protective domain	.23**	.06	.11-.35
Summary risk rating	.71*	.06	.57-.86
Unstructured clinical judgment	.55	.09	.38-.72

Note. SAVRY = Structured Assessment of Violence Risk in Youth; AUC = area under the curve; SE = standard error; CI = confidence interval.

* $p < .05$. ** $p < .01$. *** $p < .001$, two-tailed.

TABLE 5: Predictive Validity of the SAVRY for Violent Reoffending for the Detention Group (N = 40, follow-up 3 years)

	<i>AUC</i>	<i>SE</i>	<i>CI 95%</i>
Historical domain	.53	.11	.32-.75
Social/contextual domain	.68	.10	.49-.87
Individual domain	.73*	.10	.53-.93
Risk total	.74*	.10	.55-.94
Protective domain	.26*	.09	.08-.44
Summary risk rating	.76**	.08	.59-.93
Unstructured clinical judgment	.51	.10	.32-.71

Note. SAVRY = Structured Assessment of Violence Risk in Youth; AUC = area under the curve; SE = standard error; CI = confidence interval.

* $p < .05$. ** $p < .01$. *** $p < .001$, two-tailed.

of our research questions, the reconviction rates for general reoffending differed significantly, $\chi^2 = 8.7$, $p < .01$: mandatory-treatment youth (27%) and detention youth (55%).

To uncover the relative contribution of different risk assessment variables to violent reoffending, Cox regression analysis was conducted. The predictor variable of unstructured

TABLE 6: SAVRY Scores on Domains and Risk Total and Number of Cases That Were Low, Moderate, and High on Summary Risk Rating

	<i>Mandatory Treatment (N = 77)</i>		<i>Detention (N = 40)</i>	
	M	SD	M	SD
Historical	8.7	3.0	5.7***	3.2
Social/contextual	6.6	2.3	5.4*	2.5
Individual	8.1	3.7	6.1**	3.5
Risk total	23.2	6.9	17.4***	7.1
Protective	0.9	1.1	1.4*	1.2
	N	%	N	%
Summary risk rating				
Low	28	36.5%	18***	45%
Moderate	8	10.5%	6**	15%
High	41	53%	16***	40%

Note. SAVRY = Structured Assessment of Violence Risk in Youth.

* $p < .05$. ** $p < .01$. *** $p < .001$, two-tailed.

clinical judgment was entered in Step 1, the risk total score in Step 2, and the summary risk rating in Step 3. The unstructured clinical judgment accounted for a nonsignificant part of the variance (χ^2 change = .30, $p = .59$). The risk total score added a significant increment in the amount of variance explained by the unstructured judgment alone (χ^2 change = 9.5, $p = .002$), and when the summary risk rating was entered in Step 3, it produced a significant increment in the joint amount of variance explained by the unstructured judgment and risk total score (χ^2 change = 10.2, $p = .001$).

RELATIONSHIP BETWEEN UNSTRUCTURED AND STRUCTURED RISK ASSESSMENT AND SENTENCE TYPE (RESEARCH QUESTION 3)

At the time of sentencing, the juvenile judge had a forensic mental health assessment at his or her disposal, which contained an unstructured clinical risk judgment. We found that the juvenile judge's sentence showed a significant relation to the unstructured clinical risk judgment, $\chi^2(1, 2) = 27.55$, $p < .001$. High-risk youth were more likely to be sentenced to mandatory treatment and low-risk youth to detention. On the other hand, there was no relation between the structured risk assessment (SAVRY summary risk rating) and the type of sentence, $\chi^2(1, 2) = 1.92$, $p = .38$.

There was no relationship between the type of index offense and type of sentence, $\chi^2(1, 3) = 3.78$, $p = .29$. Furthermore, we also investigated the difference in risk judgment between the structured and unstructured approaches and found a significant difference, with more high-risk judgments (57% versus 49%) in the unstructured risk assessments and more low-risk judgments (15% versus 39%) in the structured risk assessments, $\chi^2(1, 4) = 24.41$, $p < .001$.

DISCUSSION

This is the first semi-prospective study in The Netherlands into the SAVRY's ability to predict violent reoffending based on forensic mental health assessments on the request of

the juvenile court. The results of this study provide support for the structured professional judgment model of risk assessment in general and for the SAVRY in particular.

We found similar AUC values for the SAVRY irrespective of type of sentence, times at risk, and type of treatment intervention. The summary risk rating, a professional judgment not made on the basis of any fixed cutoff scores, outperformed the unstructured clinical judgment in terms of predictive accuracy. The finding that structured professional judgments are better predictors than unstructured clinical judgment has been shown in research with other structured professional judgment tools as well (Borum, Bartel, & Forth, 2005; de Vogel et al., 2004; Douglas et al., 2005). We did not find empirical evidence for the hypothesis that structured clinical judgment can achieve levels of accuracy that outperform those of risk ratings based on the summation of risk scores.

With respect to the predictive accuracy of the different SAVRY domains, we found predictive values for all domains, except the historical domain. This finding is consistent with an earlier finding (Lodewijks et al., *in press*) in which the historical domain did not show predictive power for violent incidents during institutional treatment. We compared our results with those of Gretton and Abramowitz (2002), who also reported on the predictive value of the different SAVRY domains. They found a significant correlation (.22) between the historical domain and violence after release. Several studies with adult forensic psychiatric patients have reported less predictive power for the historical domain than the dynamic domains. For instance, de Vogel and de Ruiter (2006) found that a smaller number of historical items of the HCR-20 risk assessment guideline was predictive of physical violence during institutional treatment, compared to a greater number of the dynamic risk factors. Belfrage, Fransson, and Strand (2000) and Strand, Belfrage, Fransson, and Levander (1999) reported similar findings.

We were particularly interested in the relevance of risk assessment to juvenile court decision making. In our sample, we found a violent recidivism rate of 30% for the detention-sentenced youth and 14% for the mandatory-treatment-sentenced youth. These differences were significant even when corrected for time period at risk, thus indicating that detention-sentenced youths had a higher probability to recidivate. What was the reason for the differences in sentencing? Obviously, it was not the index offense, because we did not find a significant association between the type of index offense and the type of sentence. We found a significant association between the unstructured risk assessment and the type of sentence. This is not surprising because the juvenile judge asked the mental health expert for advice on recidivism risk. However, the unstructured clinical risk judgment that was used by the court on which to base its sentencing decision did not have any predictive accuracy for violent reoffending above chance.

A number of limitations to the present study should be mentioned. The first limitation relates to the semi-prospective design of the study. We could only use file information to code the SAVRY. The quality of these files differed, which may have influenced the coding. A second limitation concerns our sample, which was small. Nevertheless, we did find a number of significant differences between the samples, so apparently sample size did not lead to great loss of statistical power. Moreover, the sample was derived from only one geographical area, thereby limiting generalization. The question of generalizability of our findings to other jurisdictions deserves special attention. As already mentioned in the Method section, our sample consists largely of youth of non-European descent. A third limitation is that recidivism data were retrieved from only one source, the Judicial Documentation Register of the Ministry of Justice. As a consequence, the reconviction rate is inevitably an

underestimation of the actual recidivism rate, because not all offenders are reported, apprehended, and arrested.

It is quite common in forensic assessment for risk of future violence to be stated in terms of what might happen if no intervention would take place. In 77 out of the 117 cases, a mandatory treatment sentence was imposed, resulting in a more intensive intervention program than detention-sentenced youth would receive. The most important problem here is that risk assessment research is hampered by the clinical goals of risk assessment, as in prevention of reoffending. Thus, when forensic professionals perform SAVRY risk assessments, it is very likely that the outcome influences decisions concerning leave, entry into a semi-secure facility, or termination of treatment and that high-risk youth will not be released from the secure facility. In our study, the structured risk assessments were not available to staff in the institutions and, therefore, could not explicitly have affected their risk prevention strategies. However, it can be expected that implicit assumptions on violence risk prevention will have influenced the staff responsible for these youths. Therefore, in this study, the semi-prospective design is particularly suitable to examine the predictive validity of risk assessment instruments.

In conclusion, we offer two recommendations regarding the use of structured risk assessment instruments in forensic mental health evaluations at the request of the juvenile court. First, risk assessments should be differentiated and related to the type of sentence imposed. In the draft version of the SAVRY, there was a possibility to assess risk with or without intervention. We recommend the reintroduction of this differentiation. Second, risk assessments based on the SAVRY have a time-limited value. At least the dynamic risk items need reconsideration every half year and the protective items every year. Prospective follow-up studies longer than a year are greatly needed to evaluate whether dynamic risk factors change over time and whether this change results in a better prediction of risk. The insights that are gained from this type of study will hopefully increase mental health professionals' ability to prevent further escalation of violent juvenile offending into adulthood.

REFERENCES

- American Academy of Pediatrics. (1999). The role of the pediatrician in youth violence prevention in clinical practice and at the community level: Policy statement. *Pediatrics*, *103*, 173-181.
- Bartel, P., Borum, R., & Forth, A. (2000). *Structured Assessment for Violence Risk in Youth (SAVRY)*, consultation edition. Tampa: University of South Florida.
- 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.
- Borum, R. (2003). Managing at-risk juvenile offenders in the community. *Journal of Contemporary Criminal Justice*, *19*, 114-137.
- Borum, R., Bartel, P., & Forth, A. (2002). *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY)*, consultation edition, version 1. Tampa: University of South Florida.
- Borum, R., Bartel, P., & Forth, A. (2005). Structured Assessment of Violence Risk in Youth (SAVRY). In T. Grisso, G. Vincent, & D. Seagrave (Eds.), *Mental health screening and assessment in juvenile justice* (pp. 311-323). New York: Guilford.
- Borum, R., & Verhaagen, D. (2006). *Assessing and managing violence risk in juveniles*. New York: Guilford.
- Catchpole, R., & Gretton, H. (2003). The predictive validity of risk assessment with violent young offenders: A 1-year examination of criminal outcome. *Criminal Justice and Behavior*, *30*, 688-708.
- Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society*, *34*, 187-220.
- de Vogel, V., & de Ruiter, C. (2006). Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime, and Law*, *12*, 321-336.
- 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.

- de Vogel, V., de Ruiter, C., van Beek, D., & Mead, G. (2004). Predictive validity of the SVR-20 and the Static-99 in a Dutch sample of treated sex offenders. *Law and Human Behavior, 28*, 235-251.
- Dempster, R. J. (1998). *Prediction of sexually violent recidivism: A comparison of risk assessment instruments*. Unpublished master's thesis, Simon Fraser University, Vancouver, Canada.
- Douglas, K. S., Guy, L. S., & Weir, J. (2005). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*. Retrieved April 13, 2007, from <http://www.sfu.ca/psychology/faculty/hart/violink.htm>
- Duits, N. (2006). *De kwaliteit van onderzoek pro justitia van jongeren* [The quality of assessment in youth at the request of the court]. Unpublished doctoral dissertation, Van Gorcum, Assen, The Netherlands.
- Fitch, D. (2002). *Risk assessment on 82 high-risk Native American youth*. Unpublished doctoral dissertation, University of Texas-Clear Lake. Retrieved January 20, 2006, from http://www.fmhi.usf.edu/mhlp/savry/SAVRY_Research.htm
- Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York: John Wiley.
- Gretton, H., & Abramowitz, C. (2002, March). *SAVRY: Contribution of items and scales to clinical risk judgments and criminal outcomes*. Paper presented at the American Psychology and Law Society Biennial Conference, Austin, TX. Retrieved November 3, 2006, from http://www.fmhi.usf.edu/mhlp/savry/SAVRY_Research.htm
- Herman, S. (2005). Improving decision making in forensic child sexual abuse evaluations. *Law and Human Behavior, 29*, 87-120.
- Hogan, M. J., & Campbell, J. S. (2005). Contrasting juvenile and program-level impacts on diversion service provision: A hierarchical linear analysis. *Youth Violence and Juvenile Justice, 3*, 41-58.
- Hoge, R. (2002). Standardized instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior, 29*, 380-396.
- Hoge, R., & Andrews, D. (2002). *The Youth Level of Service/Case Management Inventory manual and scoring key*. Toronto, Canada: Multi-Health Systems.
- Hosmer, D. W., & Lemeshow, S. (1999). *Applied survival analysis*. New York: John Wiley.
- Howell, J. C., & Egle, A. (2005). Moving risk factors into developmental theories of gang membership. *Youth Violence and Juvenile Justice, 3*, 334-354.
- Koch, W. J., O'Neill, M., & Douglas, J. S. (2005). Empirical limits for the forensic assessment of PTSD litigants. *Law and Human Behavior, 29*, 121-149.
- Loeber, R., Farrington, D. P., & Waschbusch, D. A. (1998). Serious and violent juvenile offenders. In R. Loeber & D. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 13-29). Thousand Oaks, CA: Sage.
- Lodewijks, H. P. B., Doreleijers, Th. A. H., de Ruiter, C., & Borum, R. (in press). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during residential treatment. *International Journal of Law and Psychiatry*.
- Lodewijks, H. P. B., Doreleijers, Th. A. H., de Ruiter, C., & de Wit-Grouls, H. (2003). *Gestructureerde taxatie van geweldsrisico bij jongeren* [SAVRY, authorized translation into Dutch and research version]. Eefde, The Netherlands: Rentray.
- McEachran, A. (2001). *The predictive validity of the PCL:YV and the SAVRY in a population of adolescent offenders*. Unpublished master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada. Retrieved January 4, 2006, from http://www.fmhi.usf.edu/mhlp/savry/SAVRY_Research.htm
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46.
- Monahan, J. (2003). Violence risk assessment. In A. Goldstein (Ed.), *Handbook of Psychology: Forensic psychology* (Vol. 11, pp. 527-540). New York: John Wiley.
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783-792.
- Otto, R. K. (2000). Assessing and managing violence risk in outpatient settings. *Journal of Clinical Psychology, 56*, 1239-1262.
- Philipse, M. (2005). *Predicting criminal recidivism. Empirical studies and empirical practice in forensic psychiatry*. Nijmegen, The Netherlands: Radboud University.
- Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737-748.
- Schmidt, F., Hoge, R. D., & Gomes, L. (2005). Reliability and validity analyses of the Youth Level of Service/Case Management Inventory. *Criminal Justice and Behavior, 32*, 329-344.
- Sheppard, D., Rowe, W., Grant, H., & Jacobs, N. (2003). *National evaluation of the partnerships to reduce juvenile gun violence program*. Bethesda, MD: COSMOS Corporation.
- Steinberg, L., & Schwartz, R. (2000). Developmental psychology goes to court. In T. Grisso & R. Schwartz (Eds.), *Youth on trial* (pp. 9-31). Chicago: University of Chicago Press.
- 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.
- van der Heiden-Attema, N., & Wartna, B. S. J. (2000). *Recidive na verblijf in een JBI: Een vervolgstudie naar de geregistreerde criminaliteit onder jongeren uit een justitiële behandelinstelling* [Recidivism after stay in a juvenile treatment institution: A follow-up study of registered criminality in youth of a justice treatment institution]. The Hague, The Netherlands: Ministry of Justice, Wetenschappelijk Onderzoek- en Documentatie Centrum (WODC; Scientific Research and Documentation Center).

- van der Laan, P. H. (2006). Just desert and welfare: Juvenile justice in The Netherlands. In J. Junger-Tas & S. H. Decker (Eds.), *International handbook of juvenile justice* (pp. 145-170). Dordrecht, The Netherlands: Springer.
- Vida, S. (1997). *AccuROC nonparametric receiver operating characteristic analysis, version 2.5* [Computer software]. Montreal, Canada: Accumetric Corporation.
- Wartna, B. S. J., Kalidien, S. N., Tollenaar, N., & Essers, A. A. M. (2006). Strafrechtelijke recidive van jongeren uit justitiële jeugdinrichtingen: Uitstroomperiode 1997-2003 [Criminal recidivism of juveniles from juvenile justice facilities: Discharge period 1997-2003]. The Hague, The Netherlands: Ministry of Justice, WODC.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20. Assessing the risk of violence, version 2*. Vancouver, Canada: Simon Fraser University.
- Webster, C. D., Hucker, S., & Bloom, H. (2002). Transcending the actuarial versus clinical polemic in assessing risk for violence. *Criminal Justice and Behavior, 29*, 659-665.

Henny Lodewijks, Msc, is clinical psychologist and psychotherapist. He is also director of treatment to Rentray, a juvenile justice facility in The Netherlands. He is now preparing his PhD on risk assessment of violence in youth and has published articles on this topic in *Criminal Justice and Behavior*, *International Journal of Law and Psychiatry*, and *International Journal of Forensic Mental Health*.