

## **A Comparison of Modified Versions of the Static-99 and the Sex Offender Risk Appraisal Guide**

**Kevin L. Nunes,<sup>1,5</sup> Philip Firestone,<sup>1,2</sup> John M. Bradford,<sup>2</sup>  
David M. Greenberg,<sup>3</sup> and Ian Broom<sup>4</sup>**

---

*The predictive validity of 2 risk assessment instruments for sex offenders, modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide, was examined and compared in a sample of 258 adult male sex offenders. In addition, the independent contributions to the prediction of recidivism made by each instrument and by various phallometric indices were explored. Both instruments demonstrated moderate levels of predictive accuracy for sexual and violent (including sexual) recidivism. They were not significantly different in terms of their predictive accuracy for sexual or violent recidivism, nor did they contribute independently to the prediction of sexual or violent recidivism. Of the phallometric indices examined, only the pedophile index added significantly to the prediction of sexual recidivism, but not violent recidivism, above the Static-99 alone.*

---

**KEY WORDS:** sex offenders; recidivism; prediction; assessment.

---

The ability to identify sex offenders who are at high risk to reoffend provides the criminal justice system with the potential to prevent further harm to the community by these offenders. As well, some management decisions, such as civil commitment, require risk assessment. Unstructured clinical judgement has been shown to be a relatively inaccurate method of predicting sexual recidivism (Hall, 1988; Hanson & Bussière, 1998; Quinsey & Maguire, 1986). Actuarial methods have increasingly received attention in the sexual recidivism prediction literature and have demonstrated predictive accuracy superior to unstructured clinical

<sup>1</sup>School of Psychology, University of Ottawa, Ontario, Canada.

<sup>2</sup>Sexual Behaviors Clinic and Forensic Service, Royal Ottawa Hospital, Canada.

<sup>3</sup>University of Western Australia, Graylands Hospital Campus, Perth, Australia.

<sup>4</sup>Department of Psychology, Carleton University.

<sup>5</sup>To whom correspondence should be addressed at School of Psychology, University of Ottawa, 120 University Private, Ottawa, Ontario, Canada K1N 6N5; e-mail: knune008@uottawa.ca.

judgement (e.g., Hanson & Bussière, 1998). Two of the most widely used and best developed actuarial recidivism prediction scales are the Static-99 (Hanson & Thornton, 1999, 2000) and the Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, & Cormier, 1998).

The Static-99 (Hanson & Thornton, 2000) is an instrument designed to predict sexual recidivism. This measure was created by combining two prediction instruments: the Rapid Risk Assessment for Sex Offence Recidivism (RRASOR; Hanson, 1997) and the Structured Anchored Clinical Judgement—Min (SACJ-Min; Grubin, 1998). Hanson and Thornton reported that the RRASOR and the SACJ-Min appeared to be tapping into overlapping but not redundant constructs. The Static-99 is scored using relatively easily obtained information. It demonstrated superior accuracy (receiver operator characteristic [ROC] area = .71) compared to the RRASOR (ROC area = .68) and the SACJ-Min (ROC area = .67) in predicting sexual recidivism. The results were similar for violent (including sexual) recidivism with the Static-99, RRASOR, and SACJ-Min achieving areas of .69, .64, and .64, respectively. For sexual recidivism, predictive accuracy of the Static-99 was at least as high in the cross-validation sample as in the developmental samples (Hanson & Thornton, 2000). Offenders scoring in the high risk category (i.e., a score of six or higher) had a long-term sexual recidivism (15-year follow-up) rate of greater than 50%, and those scoring in the low risk category had a long-term sexual recidivism rate of 10%.

The SORAG (Quinsey et al., 1998) is a modification of the VRAG (Violent Recidivism Appraisal Guide, an instrument designed by the authors to predict violent recidivism in violent offenders), designed primarily to predict violent (including sexual) recidivism in sex offenders. It incorporates a wide range of information such as childhood history and phallometric assessment results. Compared to the Static-99, the SORAG is more comprehensive but also requires more resources to complete. The SORAG was designed using stepwise multivariate methods (Quinsey, Rice, & Harris, 1995; Rice & Harris, 1997; Rice, Harris, & Quinsey, 1990, 1991). In predicting violent (including sexual) recidivism they found that the SORAG performed slightly better with sex offenders than the VRAG (ROC area > .77; Rice & Harris, 1997). In an unpublished Master's thesis, Dempster (1998) reported ROC areas of .77 for the SORAG and .77 for the RRASOR (component of the Static-99) in predicting sexual recidivism. For violent (including sexual) recidivism, she found an area of .88 for the SORAG and .73 for the RRASOR. The SORAG was reported to be a significantly better predictor of violent (including sexual) recidivism than was the RRASOR. Although the RRASOR, which consists of four easily scored items, was just as accurate as the more comprehensive SORAG at predicting sexual recidivism, the SORAG was clearly superior in predicting the outcome for which it was designed (i.e., any violent recidivism).

In a recent study, Barbaree, Seto, Langton, and Peacock (2001) compared the predictive accuracy of a variety of risk assessment instruments, including the

Static-99 and the SORAG. Both measures showed similar levels of predictive accuracy. The ROC area for Static-99 predicting sexual recidivism was .70, and it was .70 for serious recidivism (sexual and nonsexual violence). For the SORAG, the ROC areas were .70 for sexual recidivism and .73 for serious recidivism. It was interesting to note that Barbaree et al. (2001) found that the RRASOR (Hanson, 1997) achieved a high level of accuracy in predicting sexual recidivism (ROC area of .77).

Given the importance of cross-validation of actuarial instruments (cf. Hanson, 1998), this study examined the predictive accuracy of the Static-99 and the SORAG with a new sample of generally low-risk sex offenders. In addition, the study examined whether the two instruments contributed independently to the prediction of recidivism. In light of the greater comprehensiveness of the SORAG relative to the Static-99, it was expected that the SORAG would contribute additional information to the prediction of recidivism above the Static-99 alone. An additional purpose of this study was to examine whether phallometrically assessed deviant arousal contributed additional information to the prediction of recidivism above the Static-99 alone. The SORAG incorporates phallometric assessment whereas the Static-99 does not. Hanson and Thornton (2000) suggested that the inclusion of other predictors in addition to the Static-99 items may improve predictive accuracy. In light of the strong association found to exist between phallometrically assessed arousal to children and sexual recidivism (Hanson & Bussière, 1998), predictive accuracy may be increased by considering phallometric assessment information in addition to the Static-99 alone.

## METHOD

### Participants

All participants were assessed at the Royal Ottawa Hospital, Sexual Behaviours Clinic, between 1983 and 1993. The 258 participants were males, 18 years of age or older at the time of their index offense, and had all been convicted of a hands-on sexual offense against an adult or a child (i.e., under the age of 16 at the time of the offense). The total sample of 258 sex offenders consisted of 138 incest offenders (53.5%), 107 extrafamilial child molesters (41.5%), and 13 rapists (5.0%). The majority of the participants were assessed just prior to or just after their court appearance or sentencing.

### Measures

#### *Static-99* (Hanson & Thornton, 1999)

The Static-99 consists of 10 items: (1) prior sex offenses, (2) prior sentencing dates, (3) noncontact sex offense convictions, (4) nonsexually violent index

offense convictions, (5) prior nonsexual violent convictions, (6) unrelated victim, (7) stranger victim, (8) male victim, (9) young age at time first at risk for recidivism, and (10) never married. Static-99 scores range from 0 to 12 and can be grouped into four risk categories: low (0, 1), medium-low (2, 3), medium-high (4, 5), and high (6 and up).

*Sex Offender Risk Appraisal Guide* (Quinsey et al., 1998)

The SORAG consists of 14 items: (1) lived with biological parents to age 16, (2) elementary school maladjustment, (3) history of alcohol problems (parental alcoholism, teenage alcohol problem, adult alcohol problem, alcohol involved in a prior offense, alcohol involved in index offense), (4) never married, (5) non-violent criminal history score (calculated from the Cormier-Lang system; Quinsey et al., 1998), (6) violent criminal history score (calculated from the Cormier-Lang system), (7) prior hands-on sex offense convictions, (8) female victims under 14 years old exclusively (reverse scored), (9) failure on prior conditional release, (10) young age at index offense, (11) meets *DSM-III* (American Psychiatric Association, 1980) criteria for a personality disorder, (12) meets *DSM-III* criteria for schizophrenia (reverse scored), (13) any deviant phallometric test results, and (14) Psychopathy Checklist-Revised (PCL-R; Hare, 1991) score.

*Psychopathy Checklist – Revised* (PCL-R; Hare, 1991)

The Psychopathy Checklist – Revised (PCL-R) consists of 20 clinical rating scales designed to assess behaviors and personality characteristics considered fundamental to psychopathy (Hare, 1991). Rigorous testing has indicated that the PCL-R is a psychometrically sound instrument (Hare, Forth, & Strachan, 1992). Valid PCL-R ratings can be made on the basis of high quality archival information (Harris, Rice, & Quinsey, 1994; Quinsey et al., 1995). The existence of two factors has been replicated using various samples: (1) the degree of personality, interpersonal, and affective traits deemed significant to the construct of psychopathy and (2) the degree of antisocial behaviour and an unstable, corrupted lifestyle (Hare et al., 1990). Previous studies have found the interrater reliability and internal consistency of both factors to be high despite the small number of items per factors (Hare, 1991; Hare et al., 1990).

*Phallometric Assessment*

Changes in penile circumference in response to audio/visual stimuli were measured by means of an Indium–Gallium strain gauge and were monitored by a Farrell Instruments CAT200. These data were then processed in an IBM compatible computer for storage and printout.

*Stimuli Presentation.* The order of stimulus presentation, held constant for all participants, was computer controlled using MPV-Forth version 3.05 software provided by Farrell Instruments. Participants were presented with one or more of three series of audiotapes, according to the nature of the participants' sexual offense. The audiotapes consisted of 120-s vignettes that described sexual activities varying in age, gender, and degree of consent, coercion, and/or violence portrayed (Abel, Blanchard, & Barlow, 1981). Each participant was presented with a full set containing one vignette from each category following instructions to allow normal arousal to occur. The female child series consisted of descriptions of sexual activity with a female partner/victim for eight categories. The male child series consisted of eight corresponding vignettes involving a male partner/victim, but they also included one scenario involving an adult female partner. For each of the female child and male child series, two equivalent scenarios for each category were included. The categories were (a) child initiates, (b) child mutual, (c) nonphysical coercion of child, (d) physical coercion of child, (e) sadistic sex with child, (f) nonsexual assault of child, (g) consenting sex with female adult, and (h) sex with female child relative (incest). The audiotape series used to identify sexual attraction to rape included two scenarios of 2-min duration for each of three categories: (a) consenting sex with adult female, (b) rape of adult female, and (c) nonsexual assault of adult female.

*Scoring.* The Pedophile Index (PI) was computed by dividing the highest response to the child initiates or child mutual stimulus by the highest responses to an adult consenting stimulus. Adequate discriminant validity for the PI (utilizing the same stimuli as this study) has been demonstrated by its ability to differentiate child molesters and nonoffenders significantly above chance levels of accuracy (Firestone, Bradford, Greenberg, & Nunes, 2000; Nunes et al., 1999). In terms of predictive validity, however, PI scores did not differ significantly between recidivist (sexual or violent) and nonrecidivist extrafamilial child molesters (Firestone, Bradford, McCoy, et al., 2000) or incest offenders (Firestone et al., 1999).

The Pedophile Assault Index (PAI) was computed by dividing the highest response to an assault stimulus involving a child victim (nonphysical coercion of child, physical coercion of child, sadistic sex with child, or nonsexual assault of child) by the highest response of the child initiates or child mutual stimulus. Adequate discriminant validity for the PAI (utilizing the same stimuli as this study) has been demonstrated in past studies finding significant differences between homicidal sex offenders and nonoffenders (Firestone, Bradford, Greenberg, & Larose, 1998; Firestone, Bradford, McCoy, et al., 2000). Adequate predictive validity was demonstrated for the PAI by a study that found significant differences between recidivist (sexual or violent) and nonrecidivist extrafamilial child molesters (Firestone, Bradford, McCoy, et al., 2000). However, PAI scores did not differ significantly between recidivist (sexual or violent) and nonrecidivist incest offenders (Firestone et al., 1999).

The Rape Index (RI) was computed by dividing the response to the rape stimulus by the response to the adult consenting stimulus. The Assault Index (AI) was computed by dividing the response to a nonsexual assault stimulus by the response to the adult consenting stimulus. In terms of predictive validity, RI and AI scores (utilizing the same stimuli as this study) have not been found to differ significantly between recidivist (sexual or violent) and nonrecidivist rapists (Firestone Bradford, McCoy, et al., 1998), extrafamilial child molesters (Firestone, Bradford, McCoy, et al., 2000), or incest offenders (Firestone et al., 1999). For all indices, responses in the numerator that were less than five resulted in an index of zero. Responses in the denominator that were less than five were coded as five, and the indices were then calculated.

## Procedure

### *Scoring the Static-99 and SORAG*

The instruments were scored blind to recidivism outcome. As the inclusion of a large number of offenders with missing data may not permit a fair evaluation of the optimal performance of risk assessment instruments (e.g., Rice, 2000), an effort was made to keep missing data to a minimum in this study. Specifically, only offenders with complete data for all but one (Item 3: "noncontact sex offense convictions") of the 10 Static-99 items were considered. On the SORAG, all offenders had complete information except for one item which was completely unavailable (Item 2: "elementary school problems") and a second item for which approximately half of the required information was available (Item 3: "history of alcohol problems"). More detailed information regarding these missing items is provided below. Scoring of the Static-99 followed the coding guidelines outlined in Hanson and Thornton (1999). Scoring of the SORAG followed the coding guidelines outlined in Quinsey et al. (1998). Deviations from the coding guidelines of the respective instruments are noted below. Because of the few missing variables and deviations from the coding guidelines in this study, the instruments would be most accurately described as a *modified* Static-99 and *modified* SORAG. Although the adjective "modified" may not always be used (to avoid awkwardness), any reference to the utilization of either instrument with the present sample involves the modified versions.

*Demographic Variables.* For both the Static-99 and the SORAG, in this study the "never married" item was scored based on whether the offender reported that he had ever been married or common-law regardless of the amount of time he and his partner lived together. Age at the time of the initial assessment was used to score the "age" item of both measures. The SORAG item "lived with biological parents to age 16" was scored based on whether the offender reported that he was placed outside the family home prior to age 16. The SORAG "alcohol problems" item was scored based on information about whether the offender reported a history

of alcohol dependency, family alcoholism, and that his offense was influenced by alcohol or drugs. Other information required in scoring the item concerning the "history of alcohol problems" item in the SORAG was not available for any offenders in the present sample (i.e., teenage alcohol problem and whether alcohol was involved in a prior offense). In addition, information on elementary school maladjustment was not available, and therefore the corresponding SORAG item was scored as zero for every offender in the present sample as this is the procedure suggested by the authors (Quinsey et al., 1998).

*Psychopathy.* In the present investigation, the PCL-R was completed from descriptive material contained in institutional files by two research assistants. A random sample of 100 clinic files was independently rated by each researcher, resulting in satisfactory interrater reliability ( $r = .88, p < .0001$ ). Prorated PCL-R total scores were available for all offenders in the present sample and were used to score the SORAG "psychopathy checklist score" item.

*DSM-III Diagnoses.* Diagnoses of personality disorder and schizophrenia made during the initial assessment according to *DSM-III* criteria were used to score the corresponding items of the SORAG.

*Deviant Sexual Arousal.* A score equal to or greater than 1.0 on any of the phallometric indices (i.e., Pedophile, Pedophile Assault, Rape, or Assault indices) was considered indicative of deviant sexual preferences in coding the SORAG "phallometric results" item.

*Criminal Offense History.* Offense information (i.e., prior and index offenses, conditional release failure, prior sentencing dates) was gathered from the Canadian Police Information Center (CPIC) at the Ottawa Police Station, a national database of criminal arrests and convictions including INTERPOL reports from the Royal Canadian Mounted Police. For the present sample, a distinction between formal charges that did not result in conviction and convictions was not made. The sum of these charges and convictions was used in scoring the Static-99 and SORAG criminal history items. As it was considered by the present authors to be unclear whether an offense was a noncontact sex offense or not based on the CPIC files alone and the very low frequency of official charges for noncontact sex offenses in our sample, a zero was assigned to every offender for this Static-99 item. Sex, age, number of victims, and relationship to victim (i.e., stranger, acquaintance, relative) were gathered from self-report data and used to score the corresponding Static-99 and SORAG items. Item 8 of the SORAG ("female victims under 14 years") was scored as *yes* if the offender had only female victims under age 16, as this was the closest approximation of the item available for the present sample.

### *Recidivism*

For an offender to be considered eligible to reoffend, he must have been free to commit a crime; that is, he could not have been incarcerated or in secure custody for reasons of mental illness. Recidivism information was gathered from the CPIC

records. A cumulative hierarchy in which each additional category subsumes that of the previous category was adopted to account for plea bargaining distortion and to allow comparisons with previous recidivism research (e.g., Proulx et al., 1997; Rice et al., 1991). Sexual recidivism was defined as any charge or conviction for a sexual offense after the index offense. Violent recidivism included any charge or conviction for nonsexually violent and/or sexual offenses.

### *Data Analysis*

In this study, predictive accuracy was measured using the ROC curve. The area under the ROC curve (AUC) is an improvement over other measures of predictive accuracy because it is not affected by base rates (i.e., the proportion of offenders who actually sexually recidivate), selection ratios (i.e., the proportion of offenders predicted to sexually recidivate), or both (Swets, 1986). For this reason ROC is an ideal measure for the accuracy of predictions of sexual recidivism which typically has a low base rate. The ROC curve plots the hit rate (sensitivity) against the false alarm rate ( $1 - \text{specificity}$ ) at each risk scale score. The area under the curve can range from .50 to 1.0 where .50 indicates chance level of accuracy in prediction and 1.0 indicates perfect accuracy in prediction. The AUC is a good way to describe an ROC (Swets & Pickett, 1982). The AUC can be interpreted, like the common language effect size (McGraw & Wong, 1992), as the probability that a randomly selected recidivist would have a higher risk score than a randomly selected nonrecidivist. Rice and Harris (1995) reanalyzed data from an earlier study (Harris, Rice, & Quinsey, 1993) using ROC to evaluate the predictive accuracy of the VRAG in predicting violent recidivism. They demonstrated that ROC is unaffected by varying base rates and selection ratios. In this study, data was analyzed using ROCKIT version 0.9.1 (Metz, 1998). Raw scores for the Static-99 and the SORAG were entered in the ROC analyses.

A series of sequential logistic regressions were performed to determine whether the Static-99 and SORAG contributed uniquely to the prediction of sexual and violent (including sexual) recidivism. Additional sequential logistic regressions addressed whether the Static-99 and the phallometric indices contributed uniquely to the prediction of sexual and violent (including sexual) recidivism. Specifically, the Static-99 was entered first followed by one of the following indices: the PI, PAI, RI, AI, and the highest Phallometric Index (same variable used to code the SORAG phallometric item).

## **RESULTS**

The average age at time of assessment was 38.2 years old ( $SD = 11.0$ , ranging from 18 to 73). Offenders' mean number of prior sexual charges was



**Table I.** Recidivism Rates for Static-99 Risk Levels for Present Sample and Hanson and Thornton's Sample (Hanson & Thornton, 1999)

Static-99 risk level	Sample size		Sexual recidivism		Violent recidivism	
	Present sample	Hanson & Thornton	Present sample (%)	Hanson & Thornton (%)	Present sample (%)	Hanson & Thornton (%)
Low (0, 1)	130 (50.4%)	257 (23.7%)	4.6	5.6	7.7	9.2
Medium-low (2, 3)	77 (29.8%)	410 (37.7%)	11.7	10.5	19.5	19.5
Medium-high (4, 5)	41 (15.9%)	290 (26.7%)	12.2	28.4	22.0	38.1
High (6 plus)	10 (3.9%)	129 (11.9%)	30.0	39.0	40.0	44.0
Total	258 (100%)	1086 (100%)	8.9	18.0	14.7	25.0

*Note.* The recidivism rates reported here from Hanson and Thornton (1999) are for a 5-year follow-up period.

0.31 ( $SD = 1.25$ , ranging from 0 to 15), mean number of prior nonsexually violent charges was 0.41 ( $SD = 1.29$ , ranging from 0 to 12), and mean number of prior total charges (any) was 5.7 ( $SD = 6.7$ , ranging from 0 to 52). The average follow-up time was 7.3 years ( $SD = 3.5$ ; ranging from a low of 0.80 to 25 years). The recidivism rates were 8.9% (23/258) for sexual recidivism and 14.7% (38/258) for violent recidivism. The mean score for the Static-99 was 1.91 ( $SD = 1.85$ ; ranging from 0 to 8) and for the SORAG was  $-3.19$  ( $SD = 10.26$ ; ranging from  $-19$  to 28). The low rates of recidivism and low scores on the Static-99 and SORAG reflect the relatively low risk of the present sample. The Static-99 and SORAG were significantly correlated with each other ( $r = .72$ ,  $p < .001$ ).

Presented in Table I are the distributions of Static-99 scores and recidivism rates for both the sample examined in this study and the sample studied by Hanson and Thornton (1999). As can be seen from the table, the present sample tended to be lower risk than Hanson and Thornton's sample (Hanson & Thornton, 1999). Half of the present sample fell in the lowest risk category whereas less than one quarter of Hanson and Thornton's sample fell in this category. The relatively low risk level of the present sample was also reflected in the low sexual and violent recidivism rates compared to the Hanson and Thornton's sample. In spite of these differences, both samples generally showed a linear increase in recidivism rates (both sexual and violent) as a function of their Static-99 risk categories.

Presented in Table II are the distribution of SORAG scores and recidivism rates for both the present sample and the sample used by Quinsey et al. (1998). As indicated, the present sample tended to be lower risk than Quinsey et al.'s sample (Quinsey et al., 1998). Over 50% of the present sample fell into the two lowest SORAG risk categories whereas only 15.3% of Quinsey et al.'s sample fell in these categories. Quinsey et al. did not distinguish between sexual and violent recidivism as the SORAG was designed primarily to predict any violent reoffending (sexual or nonsexual) in sex offenders, and therefore the sexual recidivism rates

**Table II.** Recidivism Rates for SORAG Risk Levels for Present Sample and Quinsey et al.'s Sample (Quinsey et al., 1998)

SORAG category	Sample size		Sexual recidivism		Violent recidivism	
	Present sample	Quinsey et al.	Present sample (%)	Quinsey et al.	Present sample (%)	Quinsey et al. (%)
1	87 (33.7%)	15 (5.2%)	5.7	NA	8.0	7.0
2	52 (20.2%)	29 (10.1%)	3.8	NA	7.7	15.0
3	46 (17.8%)	44 (15.3%)	10.9	NA	15.2	23.0
4	37 (14.3%)	58 (20.1%)	13.5	NA	29.7	39.0
5	18 (7.0%)	49 (17.0%)	22.2	NA	38.9	45.0
6	12 (4.7%)	41 (14.0%)	0	NA	0	58.0
7	4 (1.6%)	29 (10.1%)	25.0	NA	25.0	58.0
8	2 (0.8%)	20 (6.9%)	50.0	NA	50.0	75.0
9	0 (0%)	3 (1.0%)	—	NA	—	100.0
Total	258 (100%)	288 (100%)	8.9	NA	14.7	41.3

*Note.* The rates reported here from Quinsey et al. (1998) are for a 7-year follow-up period. NA = Sexual recidivism rates were not available for the Quinsey et al. (1998) SORAG data as their primary concern was with violent recidivism of any type. (Dash indicates empty cell.)

are not available for comparison with the present sample. Comparing the violent recidivism rates for the present sample and Quinsey et al.'s sample as a function of SORAG risk category, the patterns appear to be slightly different. For the Quinsey et al. sample, the progression is generally linear where recidivism tends to increase as risk category increases. This linear progression was not as apparent in the present sample. For example, 8% of offenders scoring in the lowest (first) risk bin recidivated whereas none of the offenders scoring in the sixth risk bin recidivated. A similar pattern was found for sexual recidivism. However, discrepancies between the present sample and Quinsey et al.'s results may be due to the very low number of offenders in the present sample scoring in the upper SORAG bins (i.e., inadequate number of high-risk offenders), rather than due to any inadequacy in the SORAG itself. Furthermore, the discrepancy from Quinsey et al.'s findings may have also resulted, at least in part, from the modifications made to the SORAG in the present study.

### Sexual Recidivism

ROC analyses were performed to assess the predictive validity of the Static-99 and the SORAG. For sexual recidivism, the Static-99 yielded an area under the curve (AUC) of .70 ( $SE = .05$ ,  $CI = .60-.79$ ) and the SORAG yielded an AUC of .65 ( $SE = .06$ ,  $CI = .52-.76$ ). The corresponding correlation coefficients were .18 and .17 for the Static-99 and SORAG, respectively. To assess the relative predictive accuracy of the Static-99 and SORAG, their respective AUCs were compared. The difference was nonsignificant ( $Z = .97$ ,  $p > .30$ ) indicating that performance was similar for both measures.

**Table III.** Prediction of Sexual Recidivism From Static-99 and SORAG

Scale	<i>B</i>	<i>SE B</i>	Wald
Step 1			
Static-99	0.29	.11	7.56**
Step 2			
Static-99	0.18	.15	1.35
SORAG	0.03	.03	1.04

Note.  $\chi^2(1) = 7.30$  for Step 1 ( $p < .01$ );  $\chi^2(1) = 1.02$  for Step 2 ( $p > .10$ );  $N = 258$ ; *SE* = Standard error.

\*\* $p < .01$ .

Sequential logistic regression analyses were performed to determine whether the Static-99 and SORAG, and whether the Static-99 and the phallometric indices, contributed uniquely to the prediction of sexual recidivism. In the first sequential logistic regression, Static-99 scores were entered first followed by SORAG scores. As can be seen in Table III, inclusion of both instruments did not significantly improve prediction of sexual recidivism over the Static-99 alone. In the second sequential logistic regression, Static-99 scores were entered first followed by the highest phallometric index score of the pedophile, pedophile assault, rape, and assault indices (i.e., same value used to code the SORAG phallometric item). As can be seen in Table IV, the addition of the highest phallometric index did not significantly improve prediction of sexual recidivism over the Static-99 alone. Subsequent sequential logistic regressions addressed whether each phallometric index, individually, increased prediction of sexual recidivism above the Static-99 alone. As presented in Table V, the pedophile index significantly contributed to prediction above that achieved by the Static-99 alone. In contrast, the pedophile assault, rape, and assault indices considered individually in separate sequential logistic regressions, did not contribute uniquely to the prediction of sexual recidivism beyond the Static-99 alone.

**Table IV.** Prediction of Sexual Recidivism From Static-99 and Highest Phallometric Index (of Pedophile, Pedophile Assault, Rape, or Assault Indices)

Scale	<i>B</i>	<i>SE B</i>	Wald
Step 1			
Static-99	0.29	.11	7.56**
Step 2			
Static-99	0.29	.11	7.24**
Highest Phallometric Index	0.11	.12	0.83

Note.  $\chi^2(1) = 7.30$  for Step 1 ( $p < .01$ );  $\chi^2(1) = 0.72$  for Step 2 ( $p > .10$ );  $N = 258$ ; *SE* = Standard error.

\*\* $p < .01$ .

**Table V.** Prediction of Sexual Recidivism From Static-99 and Pedophile Index (Phallometric)

Scale	<i>B</i>	<i>SE B</i>	Wald
Step 1			
Static-99	0.31	.11	8.31**
Step 2			
Static-99	0.32	.11	7.67**
Pedophile Index	0.39	.16	5.96*

Note.  $\chi^2(1) = 7.97$  for Step 1 ( $p < .01$ );  $\chi^2(1) = 5.48$  for Step 2 ( $p < .05$ );  $N = 244$ ; *SE* = Standard error.

\* $p < .05$ . \*\* $p < .01$ .

### Violent (Including Sexual) Recidivism

For violent (including sexual) recidivism, both the Static-99 and the SORAG yielded an AUC of .69 ( $SE = .04$ ,  $CI = .60-.77$ ). The corresponding correlation coefficient was .23 for both the Static-99 and SORAG. As is evident from visual inspection of these results, the difference between the AUCs of the respective measures was nonsignificant ( $Z = -.04$ ,  $p > .90$ ).

Sequential logistic regressions were performed to determine whether the Static-99 and SORAG, and whether the Static-99 and the phallometric indices, contributed uniquely to the prediction of violent (including sexual) recidivism. In the first sequential logistic regression, Static-99 scores were entered first followed by SORAG scores. As can be seen in Table VI, inclusion of both instruments did not significantly improve prediction over the Static-99 alone. In the second sequential logistic regression, Static-99 scores were entered first followed by the highest phallometric index score of the pedophile, pedophile assault, rape, and assault indices (i.e., same value used to code the SORAG phallometric item). The addition of the highest phallometric index did not significantly improve prediction over the Static-99 alone. Subsequent sequential logistic regressions addressed

**Table VI.** Prediction of Violent Recidivism From Static-99 and SORAG

Scale	<i>B</i>	<i>SE B</i>	Wald
Step 1			
Static-99	0.32	.09	12.21***
Step 2			
Static-99	0.18	.13	1.99
SORAG	0.03	.02	2.11

Note.  $\chi^2(1) = 12.31$  for Step 1 ( $p < .001$ );  $\chi^2(1) = 2.08$  for Step 2 ( $p > .10$ );  $N = 258$ ; *SE* = Standard error.

\*\*\* $p < .001$ .

whether the PI, PAI, RI, and AI, individually, increased prediction of violent recidivism above the Static-99 alone. None of the individual phallometric indices contributed uniquely to the prediction of violent recidivism beyond that achieved by the Static-99 alone.

## DISCUSSION

The primary purpose of this study was to examine and compare the accuracy of the Static-99 and SORAG in predicting sexual and violent (including sexual) recidivism in a new sample of sex offenders. Both the Static-99 and SORAG achieved moderate predictive accuracy with the present sample for both sexual and violent recidivism. Despite the relatively low risk of the present sample (53.5% were incest offenders) and the modifications made to the instruments, the Static-99 and the SORAG predicted sexual recidivism in this study (ROC areas = .70 and .65, respectively) with levels of accuracy similar to other research (Barbaree et al., 2001; Hanson & Thornton, 2000). Similarly, for violent (including sexual) recidivism, performance of the Static-99 and the SORAG in this study (ROC area = .69 for both instruments) is comparable to their performance reported by Hanson and Thornton (1999; ROC area = .69 for Static-99) and Barbaree et al. (2001; ROC area = .73 for SORAG and .70 for Static-99).

A second purpose of this study was to determine whether the Static-99 and SORAG contributed independently to the prediction of recidivism. Results indicated that the instruments did not contribute uniquely to the prediction of sexual or violent recidivism, suggesting that the instruments may provide redundant information. The current research also examined whether phallometric assessment information contributed independently to the prediction of recidivism once Static-99 scores were taken into account. When the phallometric indices were added to the Static-99, only the PI contributed independently to the prediction of sexual recidivism. This suggests that accuracy in the prediction of sexual recidivism could be increased by combining the Static-99 and PI, above the accuracy achieved by the Static-99 alone. The independent contribution of the PI is not surprising, given the strength demonstrated by this predictor in the general literature (Hanson & Bussière, 1998). In contrast, the PI did not contribute uniquely to the prediction of violent recidivism beyond the Static-99. In addition, the PAI, RI, AI, and a score consisting of the maximum value within all the indices mentioned above (i.e., PI, PAI, RI, and AI), did not contribute independently to the prediction of sexual recidivism or violent recidivism beyond the contribution of the Static-99. The null results for the RI are consistent with the generally weak relationship found between phallometric assessment of arousal to rape and sexual recidivism (Hanson & Bussière, 1998). Perhaps if the SORAG included only arousal to children rather than "any deviant arousal," it would have demonstrated greater predictive accuracy in this sample.

Both Hanson and Thornton (2000) and Quinsey et al. (1998) found a generally linear increase in recidivism rates as a function of risk category for the Static-99 and the SORAG, respectively. In the present sample, a similar pattern was seen for the Static-99 but was not as apparent for the SORAG. This may reflect a real discrepancy between SORAG-predicted risk level and actual risk level in the current sample. However, only 11, 4, 2, and 0 offenders in the present sample fell in SORAG bins 6, 7, 8, and 9, respectively. The low number of offenders falling into the upper SORAG risk bins and the modifications made to the instrument in this study preclude strong conclusions regarding the validity of the risk bins. Nonetheless, the discrepancy is consistent with research on the VRAG that has also failed to replicate the linear increase in recidivism as a function of risk bin (Douglas, Hart, Dempster, & Lyon, 1999).

A limitation of this study is the presence of missing data and slight modifications of coding procedures. It has been argued that the fairest evaluations of the Static-99 and SORAG would involve no modifications and no missing data (e.g., Rice, 2000). In the present investigation, data was missing for only one item out of 10 for the Static-99 and two items out of 14 for the SORAG. This is fairly common in these types of studies (e.g., Barbaree et al., 2001; Hanson & Thornton, 1999). Nevertheless, it must be acknowledged that the results might have differed had the instruments been scored exactly as intended by their authors. Modifications and missing data can dilute the predictive power of an instrument. Modifying the SORAG item "female victims under 14 years of age" to "female victim under 16 years of age," for example, may have reduced the predictive validity of the item. Some critical dimension of this predictor, such as the absence of secondary sex characteristics, may have been lost by the modification. In addition, because of the SORAG's comprehensiveness, it tends to be at an unfair disadvantage relative to more easily scored instruments in terms of modifications that must be made and missing data. For example, more items were modified and missing for the SORAG than for the Static-99 in this study as well as in other research (Barbaree et al., 2001). A second limitation is that interrater reliability could not be assessed in this study as many of the variables in both instruments were recoded from previously entered data.

In spite of the moderate predictive validity of actuarial prediction instruments demonstrated in this study as well as in other research, arguments have been made against overreliance on these instruments in forensic assessments (Boer, Hart, Kropp, & Webster, 1997; Boer, Wilson, Gauthier, & Hart, 1997; Buchanan, 1999; Grubin, 1997, 1999; Grubin & Wingate, 1996; Rodgers, 2000). Current actuarial prediction instruments for sex offenders generally consider primarily static variables, and, as such, they have been criticized for not providing information regarding treatment targets, monitoring change, and for not taking into account case-specific factors and potential interactions between risk factors (e.g., psychopathy and deviant sexual arousal). However, recent work has begun to address some of these criticisms (e.g., Hanson & Harris, 2000).

Future research may provide the fairest evaluations of these instruments using complete data for all items with no modifications and a higher proportion of high-risk offenders. In addition, predictive accuracy based on a combination of static variables such as those included in the Static-99 or SORAG and dynamic variables such as those considered in the Sex Offender Need Assessment Rating (SONAR; Hanson & Harris, 2000), for example, may potentially provide more refined predictions, indicate treatment targets, and provide a measure of change in risk level. Finally, structured clinical judgement schemes such as the Sexual Violence Risk – 20 Scale (SVR-20; Boer, Hart, et al., 1997; Boer, Wilson, et al., 1997) should be further examined as they may provide a useful alternative to current actuarial instruments (Dempster, 1998).

The results of the present investigation support the continued use of the Static-99 and the SORAG in sex offender risk assessment. Both demonstrated similar levels of predictive validity for both sexual and violent recidivism and appeared to provide overlapping information regarding risk for recidivism. To the extent that the modifications in this study did not differentially affect the accuracy of the instruments, the Static-99, in light of its relative ease of administration, may be the more sensible choice for evaluators who do not have the time required for the SORAG.

## ACKNOWLEDGMENTS

We thank R. Karl Hanson and an anonymous reviewer for suggesting to evaluate the independent contribution made by each instrument. We also thank R. Karl Hanson for suggesting to evaluate the independent contribution made by phallometric data and the Static-99 through sequential logistic regression.

## REFERENCES

- Abel, G. G., Blanchard, E. B., & Barlow, D. H. (1981). Measurement of sexual arousal in several paraphilias: The effects of stimulus modality, instructional set, and stimulus content on the objective. *Behavior Research and Therapy*, 19, 25–33.
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, 28, 490–521.
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk—20: Professional guidelines for assessing risk of sexual violence*. Vancouver, BC: BC Institute Against Family Violence and the Mental Health, Law, and Policy Institute, Simon Fraser University.
- Boer, D. P., Wilson, R. J., Gauthier, C. M., & Hart, S. D. (1997). Assessing risk of sexual violence: Guidelines for clinical practice. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment, and treatment* (pp. 326–342). New York: Guilford.
- Buchanan, A. (1999). Risk and dangerousness. *Psychological Medicine*, 29, 465–473.

- Dempster, R. J. (1998). *Prediction of sexual violent recidivism: A comparison of risk assessment instruments*. Unpublished master's thesis, Simon Fraser University.
- Douglas, K. S., Hart, S. D., Dempster, R. J., & Lyon, D. R. (1999, July). *The violence risk appraisal guide: An attempt at validation in a sample of maximum security forensic psychiatric patients*. Poster session presented at the Joint American Psychology-Law Society (APA Div. 41)/European Association of Psychology and Law Conference, Dublin, Ireland.
- Firestone, P., Bradford, J. M., Greenberg, D. M., & Larose, M. R. (1998). Homicidal sex offenders: Psychological, phallometric, and diagnostic features. *Journal of the American Academy of Psychiatry and Law*, 26, 537-552.
- Firestone, P., Bradford, J. M., Greenberg, D. M., & Nunes, K. L. (2000). Differentiation of homicidal child molesters, nonhomicidal child molesters, and nonoffenders by phallometry. *American Journal of Psychiatry*, 157, 1847-1850.
- Firestone, P., Bradford, J. M., McCoy, M., Greenberg, D. M., Curry, S., & Larose, M. R. (1998). Recidivism in convicted rapists. *Journal of the American Academy of Psychiatry and Law*, 26, 185-200.
- Firestone, P., Bradford, J. M., McCoy, M., Greenberg, D. M., Curry, S., & Larose, M. R. (2000). Prediction of recidivism in extrafamilial child molesters based on court related assessments. *Sexual Abuse: A Journal of Research and Treatment*, 12, 203-221.
- Firestone, P., Bradford, J. M., McCoy, M., Greenberg, D. M., Larose, M. R., & Curry, S. (1999). Prediction of recidivism in incest offenders. *Journal of Interpersonal Violence*, 14, 511-531.
- Grubin, D. (1997). Predictors of risk in serious sex offenders. *British Journal of Psychiatry*, 170, 17-21.
- Grubin, D. (1998). *Sex offending against children: Understanding the risk*. Police Research Series Paper 99. London: Home Office.
- Grubin, D. (1999). Actuarial and clinical assessment of risk in sex offenders. *Journal of Interpersonal Violence*, 14, 331-343.
- Grubin, D., & Wingate, S. (1996). Sexual offence recidivism: Prediction versus understanding. *Criminal Behaviour and Mental Health*, 6, 349-359.
- Hall, G. C. N. (1988). Criminal behaviour as a function of clinical and actuarial variables in a sexual offender population. *Journal of Consulting and Clinical Psychology*, 56, 773-775.
- Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism* (User Report 97-04). Ottawa, Canada: Department of the Solicitor General of Canada.
- Hanson, R. K. (1998). What do we know about sex offender risk assessment? *Psychology, Public Policy, and Law*, 4, 50-72.
- 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. (2000). *The Sex Offender Need Assessment Rating (SONAR): A method for measuring change in risk levels* (User Report 2000-01). Ottawa, Canada: Department of the Solicitor General of Canada.
- Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders* (User Report 99-02). Ottawa, Canada: Department of the Solicitor General of Canada.
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24, 119-136.
- Hare, R. D. (1991). *Manual for the revised psychopathy checklist*. Toronto, Canada: MultiHealth Systems.
- Hare, R. D., Forth, A. E., & Strachan, K. E. (1992). Psychopathy and crime across the life span. In R. D. Peters, J. McMahon, & V. L. Quinsey (Eds.), *Aggression and violence throughout the life span* (pp. 285-300). Newbury Park, CA: Sage.
- Hare, R. D., Harpur, T. J., Hakstian, A. R., Forth, A. E., Hart, S. D., & Newman, J. P. (1990). The Revised psychopathy Checklist: Reliability and factor structure. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2, 338-341.
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, 20, 315-335.
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1994). Psychopathy as a taxon: Evidence that psychopaths are a discrete class. *Journal of Consulting and Clinical Psychology*, 62, 387-397.



- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, 111, 361–365.
- Metz, C. E. (1998). ROCKIT (Version 0.9.1) [Computer software]. Chicago: University of Chicago.
- Nunes, K. L., Firestone, P., Bradford, J. M., Greenberg, D. M., Smith, D., & Serran, G. (1999, September). *The relative accuracy of auditory and visual stimuli in phallometric assessment to identify extrafamilial child molesters*. Poster session presented at the 18th annual research and treatment conference of the Association for the Treatment of Sexual Abusers, Lake Buena Vista, FL.
- Proulx, J., Pellerin, B., Paradis, Y., McKibben, A., Aubut, J., & Ouimet, M. (1997). Static and dynamic predictors of recidivism in sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 9, 7–27.
- 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., & Maguire, A. (1986). Maximum security psychiatric patients: Actuarial and clinical prediction of dangerousness. *Journal of Interpersonal Violence*, 1, 143–177.
- Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence*, 10, 85–105.
- Rice, M. E. (2000). *Assessing risk among sex offenders: Prospects for dynamic predictors*. Paper presented at the 19th annual conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.
- 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.
- Rice, M. E., Harris, G. T., & Quinsey, V. L. (1990). A followup of rapists assessed in a maximum security psychiatric facility. *Journal of Interpersonal Violence*, 5, 435–448.
- Rice, M. E., Harris, G. T., & Quinsey, V. L. (1991). Evaluation of an institution-based treatment program for child molesters. *Canadian Journal of Program Evaluation*, 6, 111–129.
- Rodgers, R. (2000). The uncritical acceptance of risk assessment in forensic practice. *Law and Human Behavior*, 24, 595–605.
- Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, 99, 100–117.
- Swets, J. A., & Pickett, R. M. (1982). *Evaluation of diagnostic systems*. New York: Academic Press.