

**RELIABILITY AND VALIDITY
STUDY
OF THE STATIC-99/RRASOR
SEX OFFENDER RISK
ASSESSMENT
INSTRUMENTS**

James Austin, Ph.D.
Johnette Peyton, M.S.
Kelly Dedel Johnson, Ph.D.

Final Report

*Submitted to the
Pennsylvania Board of Probation and Parole
January, 2003*

The Institute on Crime, Justice and Corrections
at The George Washington University
10 G Street, NE, Suite 610
Washington, DC 20002
202-408-6800

ACKNOWLEDGMENTS

The Institute on Crime, Justice and Corrections would like to acknowledge the Pennsylvania Board of Probation and Parole for its continuous efforts to develop and implement a new set of guidelines that better reflect the Board's current policies regarding parole. This study would not have been possible without the dedication and hard work of the PBPP Research and Statistics Division. Special recognition is due to Jim Alibrio, Director of Research and Statistics, and his dedicated staff who made significant contributions to the design and data collection efforts required for this study. Also, we wish to express our gratitude to the Pennsylvania Commission on Crime and Delinquency for sponsoring this project. In particular, our PCCD project manager, Lonisha Morgan, helped us to complete this project in a timely and thoughtful manner.

This study was funded by the Pennsylvania Commission on Crime and Delinquency and is in response to DESI Program Purpose Area #11. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the Pennsylvania Commission on Crime and Delinquency or the Pennsylvania Board of Probation and Parole.

EXECUTIVE SUMMARY

In late 1999, the Sex Offender Assessment Board (SOAB), for its work assessing sexual offenders, both for court ordered and PBPP cases, began use of the RRASOR, Static 99, and MNSOST-R. The PBPP, upon recommendation of the SOAB, adopted the Static 99 as its objective risk assessment instrument to assist in determining the static level of risk to re-offend sexually posed by sex offenders. The Static-99 was first developed by Canadian researchers and includes ten items that take into account the offender's prior criminal history, incidence of violence, relationship to the victim, and personal characteristics that may influence the decision of the PBPP. These variables have been shown to be statistically associated with recidivism in other jurisdictions. Within the Static-99, there are four items that compose a condensed risk assessment instrument known as the Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR).

Summary of Research Design and Methodology

In order to measure the reliability of Static-99/RRASOR, two reliability tests were completed. The first consisted of a pilot test of 41 cases that were coded by independent scorers using case file data. A second test was conducted on a larger sample of 220 cases that were drawn for the validation study.

The validation sample consisted of 550 cases that reflected prisoners convicted of sex crimes and were released either between 1994 and 1995 (N=356) or in 2000 (N=194). The characteristics of the 550 cases were compared to the total universe of all released sex offenders. This analysis found the sample to be representative of all sex releases for the two time periods.

Reliability Findings

Two reliability tests were conducted to ascertain the reliability of three sex offender risk assessment instruments. In addition to the Static-99/RRASOR, the reliability of the Minnesota Sex Offender Screening Tool-Revised (MnSOST-R) was also assessed.

The first test was a limited pilot study that consisted 41 cases, each scored twice on the three instruments. In order for the scoring of the items to be judged reliable, the two raters needed to assign the same score for at least 80 - 90 percent of the cases.

The pilot study revealed that the level of inter-rater reliability was unacceptably low for the MnSOST-R, based in large part on the information collected in the records of both the DOC and PBPP. While, the Static 99/RRASOR had higher levels of reliability, significant staff training and item refinement/clarification was required to improve the scoring of the Static 99/RRASOR; DOC and PBPP records are not a factor for RRASOR/Static 99.

The second reliability test on 220 cases showed significant improvement in the level of scoring consistency achieved across raters. However, the raters agreed in only 73 percent of the cases. The rates of reliability were much higher for cases released in the 2000 largely due to the higher quality of data available in the PBPP files.

Validation Findings

For both release cohorts, the items used in the Static-99/RRASOR instrument are associated with recidivism. Furthermore, the risk levels are also associated to recidivism. The single item that could be improved is the age factor by changing the threshold level from 25 to 45 years at the time of release.

Conclusions and Recommendations

The Static 99/RRASOR is a reliable and valid instrument for assessing sex offenders' risk to recidivate. However, more intensive and effective training is needed to insure consistent scoring across raters. It is also recommended that the age variable be adjusted on the instrument to enhance the already acceptable level of prediction associated with this assessment instrument.

I. INTRODUCTION

A. Background

In late 1999, upon recommendation by the Sexual Offenders Assessment Board (SOAB), the PBPP adopted the Static 99 as its objective risk assessment instrument to determine the appropriate post-release supervision level for sex offenders on parole. The Static-99 was first developed by Canadian researchers and includes ten items that take into account the offender's prior criminal history, incidence of violence, relationship to the victim, and personal characteristics that have relevance to the PBPP parole decision. These variables have been shown to be statistically associated with recidivism in other jurisdictions. Within the Static-99, there are four items that compose a condensed risk assessment instrument known as the Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR). The four items comprising the RRASOR largely account for the predictive ability of the Static-99.

In October 2001, The Institute on Crime, Justice and Corrections at The George Washington University and the SOAB conducted a pilot test to ascertain the reliability of three sex offender risk assessment instruments. In addition to the Static-99/RRASOR, the reliability of the Minnesota Sex Offender Screening Tool-Revised (MnSOST-R) was also assessed. The pilot study consisted of a small sample of 41 sex offenders, each scored twice on the three instruments drawn from both a 1994-95 and a year 2000 release cohort.

In order for the scoring of the items to be judged reliable, the two raters needed to assign the same score for at least 80 percent of the cases and optimally 90 percent of the cases. This was an ex-post facto scoring using case folders from both the Department of Corrections and the Parole Board. The pilot study revealed that the level of inter-rater reliability was unacceptably low for the MnSOST-R because many instrument questions could not be scored from available information. The Static-99/RRASOR had higher rates of reliability. However, significant staff training and item refinement were required to ensure the consistent application of these instruments to Pennsylvania's sex offender parole population.

Subsequently in February 2002, the SOAB initiated an ambitious course of action to remedy the identified rater reliability problems and score the targeted 550 cases in the original study design for purposes of validation study. Because of the poor reliability results and the fact that the MnSOST-R is a more difficult instrument to score, the SOAB decided to focus its efforts solely on the Static-99/RRASOR instruments. The data collection effort included a quality control program in which staff refined the operational definitions of the Static-99/RRASOR items to ensure their clarity and specificity and received intensive training in the administration of the instruments.

The Static-99 consists of 10 items, which are believed to be related to future criminal behavior. Those items noted by asterisk (*) comprise the RRASOR risk assessment instrument:

1. Prior Sex Offenses*
 - a. Charges
 - b. Convictions
2. Prior Sentencing Dates
3. Any Convictions for Non-Contact Sex Offenses
4. Index Non-Sexual Violence
5. Prior Non-Sexual Violence
6. Any Unrelated Victims*
7. Any Stranger Victims
8. Any Males Victims*
9. Young*
10. Single

Through a file review process, offenders are scored on items requiring either a yes/no response or the use of a structured scale ranging in value from 0 to 3. Based on these answers, the rater computes the total score by summing all the items and then determines the offender's overall risk level using the Static-99 scale. Offenders scoring between 0 and 1 are considered "low risk," those scoring between 2 and 3 are considered "medium-low risk," those scoring between 4 and 5 are considered "medium-high risk," and those scoring between 6 and 10 are considered to be "high risk."

B. Project Objectives

The overall goal of this study is to assess the use of the Static-99/RRASOR for its suitability for use as a risk assessment instrument for Pennsylvania's sex offender parole population. The two separate, but related, objectives of the study are to assess the level of inter-rater reliability and the level of validity of the instruments. Reliability refers to the level of consistency in scoring prisoners on the various factors used by the two instruments to determine the appropriate level of post-release supervision. There are two types of reliability, inter- and intra-rater reliability. The former involves whether two persons computing the score on the Static-99/RRASOR for the same individual reach the same rating. Intra-rater reliability refers to whether a single rater scoring the Static-99/RRASOR for the same prisoner on repeated applications will obtain the same rating. Inter-rater reliability is the focus of the current research. Reliability is an essential component of any risk assessment instrument, as a jurisdiction must be confident that the instrument will be applied consistently across offenders and staff in order to ensure its objectivity. An instrument cannot be considered valid if it is not reliable, that is, if it is not applied consistently.

Specific research questions guiding this study are as follows:

Reliability Study:

1. To what extent is the scoring of the Static-99/RRASOR reliable?
2. Which items on the Static-99/RRASOR have acceptable rates of inter-rater reliability and which ones require additional staff training or modification to improve their levels of reliability?
3. Is there consistency in the final risk levels computed across raters?

Validation Study:

1. To what extent is the Static-99/RRASOR valid?
2. Is the instrument able to distinguish distinct groups of offenders with different likelihoods of recidivating?
3. Are there adjustments that could be made to the instrument or to the cut-points on the scale that would improve its validity?

These questions will be considered in the following sections of the report.

C. Project Sampling Overview

A data file was obtained from the Pennsylvania Department of Corrections (DOC) in August 2001 that identified 846 cases released from state prisons in calendar years 1994 and 1995 with sex offense codes identified by the Sexual Offenders Assessment Board (SOAB). These cases provided the universe for a five year follow up study of which 100 cases would be targeted for an ex-post facto rater reliability study. A second data file of 277 cases identified all sex offenders released onto parole in calendar year 2000 by the PBPP that met SOAB sex offense specifications. This second release cohort was used for a one-year follow up study of which about 100 also would be studied for inter-rater reliability. Twenty-five cases from each of these two release cohort groups subsequently were selected randomly for an inter-rater reliability pilot study of which 41 were scored and used for an analysis of inter-rater reliability on three risk assessment instruments.

A pool of 1,009 cases eventually was identified as available study subjects for a validation study and a full inter-rater reliability study after subtracting the 50 cases used in the pilot study. A number of duplicate records were eliminated because cases appeared in both the 1994-95 and the 2000 release cohort; the earliest release year was the criteria for selection of remaining cases. The project design called for a sample size of 550 cases of which approximately 200 would be scored twice on each risk instrument. A stratified random selection strategy was used by PBPP research staff to identify 550 cases out of 1,009 in the subject pool, which would be representative of cases that were active or inactive (archived folders), and from the two release cohorts. Only PBPP case folders

were used to score the selected 550 cases for the validation study of which 220 were scored twice for purposes of the inter-rater reliability study.

In March 2002, the SOAB implemented an aggressive three-phase plan to score the 550 cases for validation follow up and inter-rater reliability. This required an organized labor-intensive folder pulling process by PBPP from the Pennsylvania State Archives and the Board's file room. Phase 1 required six SOAB staff to score a set of cases twice and identify quality control issues for future scoring. Phase 2 required six SOAB staff to score a second set of cases twice with six institutional staff in train-the-trainer sessions to complete the inter-rater reliability data collection. In Phase 3, the remaining cases folders were scored to complete the validation study. Two practical factors intervened during all data collection phases that required replacement of some target folders and causing cases to be dropped from the analysis. First, a number of cases were misidentified as being eligible for inclusion (e.g., they did not have a prior sex offense, or were not male—the instruments have only been validated on males). Second, a number of case files were not accessible because they were in distribution or could not be retrieved from the state archives in a timely manner. Random selection was used in replacing cases, usually by picking the next available case from master lists.

II. RELIABILITY STUDY

A. Methodology

1. Sampling Method

In order to measure the reliability of Static-99/RRASOR, it was necessary to draw a representative sample of all released sex offenders. A data file from the PBPP provided a profile of the demographic and offense characteristics of the total population of paroled sex offenders from the two cohorts released from 1994-95 and 2000. As previously discussed, the total population totaled 1,009 offenders, and a random sample of 220 cases was selected and scored twice for inter-rater reliability testing.

2. Data Collection

During Phase I, 45 cases were scored twice by two separate SOAB staff who were trained in the application of Static-99. A total of six SOAB staff members were involved in this phase of the data collection. In addition to scoring cases, these staff were also tasked with identifying and addressing quality control issues that could impede in the reliability of the risk assessment instruments. Phase II enlisted six additional staff from PBPP to score an additional 175 cases. These additional staff were selected to ensure that institutional staff, who will be using Static-99/RRASOR on the sex offender population in the future, were represented among those involved in the study. The team was divided into two groups: A and B, each of which consisted of an equal number of SOAB and PBPP staff.

After the A or B group member scored a case, it was randomly assigned to a member of the other group for the second scoring.

Also, after-the-fact quality control efforts were implemented. On a given case, if there was more than a two point difference in the total score assigned by each rater, a 'rater consultation' was held to determine the cause of the discrepancy and to identify methods by which greater consistency could be achieved. These consultations did not change the total scores that were assigned, thereby maintaining the integrity of the research methodology.

An ICJC staff member monitored the data collection process to ensure authenticity of the data collection and data entry process. For example, the scoring and consultation sessions of the PBPP and SOAB raters and the entering the scored cases into an MS Access database were observed in phase 1. Once all cases were scored and key entered, the PBPP sent an electronic copy of the data file that was then converted by ICJC staff into an SPSS database for statistical analysis.

B. Results

Descriptive statistics were used to determine if the sample was representative of all released sex offenders and to assess the level of (or percent) agreement between the two interviewers on each item, and to examine differences in the distribution across risk levels for the two raters. For the purposes of this reliability assessment, a percent agreement statistic was computed, which reflects the level of agreement between the raters in their scores on each item and the assigned risk level.

Table 1 presents compares the demographic and offense characteristics of the sample with those of the total population, and reveals no significant differences. Over 95 percent of the offenders were male. In terms of race/ethnicity, 47 percent of the sample was Black, 47 percent was white, and 5 percent was Hispanic. The largest proportion of offenders (44%) was between 30 and 39 years old. More than half of the offenders (52%) were charged with rape, and one-fifth (21%) were charged with involuntary deviant sexual intercourse. Seventy-eight percent of the offenders were released on parole and approximately 19 percent had reached the maximum date of their sentence.

Table 2 shows the level of agreement between the two raters across the ten items that comprise the Static-99/RRASOR. As discussed previously, the RRASOR includes only four items that share the exact wording of their Static-99 counterpart. Therefore, the reliability of the RRASOR can be discerned by examining the rates of agreement on those items marked with an asterisk. Only two items meet the 90 percent threshold: "Any male victims?" and "Young (under age 25)?" Two items met the 90 percent criteria while the level of agreement on the remaining eight variables was over 80 percent. This represents a marked improvement in consistency as compared to the initial pilot test conducted in October, 2001.

However, the level of agreement on the total score was only 41 percent, while the level of agreement on the assessed risk level was 73 percent. The assessed level has a certain level of built-in tolerance, that is, a range of scores will suggest the same risk level. Thus, the low level of agreement in the total score is not as disconcerting as it first seems. However, although the level of agreement for the assessed risk level is higher, the fact that in 27 percent of cases the raters assessed the same prisoner at different risk levels (and thus, requiring different levels of post-release supervision) is cause for concern.

The distribution of cases across risk levels for the first and second raters is presented in Table 3, and clearly, does not differ radically. The first raters scored 45 percent of the sample as low risk, compared 41 percent of the second raters. Similarly, the first raters scored 42 percent of the sample as medium risk, compared to 47 percent of the second raters. There was very few high-risk cases in the sample and little difference in the proportion of cases scored by the first (13%) and second (12%) rater.

However, there are significant differences in the assessed risk level of individual cases, Table 4 presents a more direct comparison of the level and direction of risk level disagreement. The shaded, diagonal boxes in the table represent those cases in which the two interviewers agreed on the prisoners' risk level, which represent 73 percent of the total sample. For the cases in which there was disagreement, in general, the raters differed by only one risk level (e.g., one scored the prisoner as low and the other scored the prisoner as medium) rather than two risk levels (e.g., low versus high). The lack of reliability in the assessed risk level makes the validation of the instrument challenging because in 27 percent of the cases, the risk level of an prisoner is questionable as it is likely that another rater would have scored the prisoner differently, even if by only one level.

In comparing the two release cohorts, the 2000 release group was found to have higher reliability rates. For Static-99 and RRASOR, the 2000 release group had two highly reliable variables: the "Prior sex offenses charges" and "Any unrelated victims" items. Also, the "Any male victims" and "Young" items were found to be considerably reliable in both release groups. This high consistency of scoring among the 2000 release cases could be due to the PBPP files providing more relevant information after implementation of Megan's law than was available at the time that the 1994, 1995 group's records were created.

C. Conclusions and Recommendations

In comparison to the pilot test, there has been a significant improvement in the level of scoring consistency achieved across raters on the Static-99/RRASOR. These improvements testify to the importance of the comprehensive training effort initiated as part of Phase I to improve the consistency and specificity in scoring. However, significant differences in total score and assessed risk level remain, which indicates an enduring need for continued training. High levels of reliability are considered essential in order to have confidence in the findings of the validation study.

More intensive and effective training is needed to insure consistent scoring across raters. One could achieve higher reliability levels if all raters were provided effective training during simultaneously instead of breaking correctional staff into groups. In addition, a system of internal monitoring should be implemented to sustain a high level of consistency. Also, it is possible that the risk level scale may need some adjustment.

Table 1 Comparison of Population and Sample Demographic and Offender Characteristics				
Characteristics	Population		Sample	
	N=1,009	100%	N=550	100%
Sex				
Male	971	96.2	550	100.0
Female	38	3.8	0	0.0
Race				
White	478	47.4	263	47.8
Black	476	47.2	257	46.7
Hispanic	50	5.0	28	5.1
Asian	3	0.3	1	0.2
Indian	2	0.2	1	0.2
Age Categories				
20-24	44	4.4	28	5.1
25-29	145	14.4	77	14.0
30-34	200	19.8	105	19.1
35-39	244	24.2	135	24.5
40-44	163	16.2	89	16.2
45-49	110	10.9	61	11.1
50-54	47	4.7	27	4.9
55+	56	5.6	28	5.1
Release Type				
Bail	6	0.6	3	0.5
County	3	0.3	2	0.4
Court	10	1.0	5	0.9
Death	21	2.1	9	1.6
Maxed out	193	19.1	102	18.5
Paroled	771	76.4	428	77.8
Other	5	0.5	1	0.2

Table 1 Comparison of Population and Sample Demographic and Offender Characteristics				
Characteristics	Population		Sample	
	N=1,009	100%	N=550	100%
Type of Offense				
Aggravated Sexual Assault	17	1.7	8	1.5
Incest	5	0.5	3	0.5
Indecent Assault	98	9.7	51	9.3
Indecent Exposure	8	0.8	5	0.9
Involuntary Deviate Sexual Intercourse	231	22.9	117	21.3
Kidnapping	33	3.3	17	3.1
Prostitution	30	3.0	15	2.7
Rape	513	50.8	286	52.0
Sex Abuse of Children	2	0.2	1	0.2
Spousal Sexual Assault	2	0.2	0	0.0
Statutory Rape	67	6.6	45	8.2
Voluntary Deviate Sexual Intercourse	3	0.3	2	0.4
<i>Source: Pennsylvania Board of Probation and Parole</i>				

Table 2. Percent Agreement Between Raters for Static-99/RRASOR	
Item	% Agreement
1. Prior sex offenses Charges** Convictions**	82.7%
2. Prior sentencing dates	87.7%
3. Any convictions for non-contact sex offense	86.4%
4. Index non-sexual violence	81.8%
5. Prior non-sexual violence	87.7%
6. Any unrelated victims**	82.7%
7. Any stranger victims	86.8%
8. Any male victims**	96.8%
9. Young**	98.6%
10. Single	81.4%
11. Total Score	40.9%
12. Risk Level	73.2%
<i>Notes: an ** is used to identify those items that comprise the RRASOR. Source: PBPP/Sexual Offenders Assessment Board, Phase II data</i>	

**Table 3.
Comparison of Static-99/RRASOR Risk Levels Across Raters of the Reliability Sample**

Static-99/RRASOR Risk Levels	Rater One		Rater Two	
	N	100%	N	100%
Low (0 through 2)	98	44.5	90	40.9
Medium (3 through 5)	93	42.3	103	46.8
High (6 and above)	29	13.2	27	12.3
TOTAL	220	100.0	220	100.0
<i>Source: PBPP/Sex Offender Assessment Board</i>				

Table 4. Cross-Tabulation of Static-99/RRASOR Risk Levels Across Raters of the Reliability Sample				
Rater Two	Rater One			TOTAL
	Low	Medium	High	
Low (0 through 2)	76	13	1	90
Medium (3 through 5)	22	69	12	103
High (6 and above)	0	11	16	27
TOTAL	98	93	29	220
% Disagreement: 27 %; % Disagreement One Level: 26 %; % Disagreement Two Levels: 0.5 %				
<i>Source: PBPP/Sexual Offenders Assessment Board</i>				

III. VALIDATION OF STATIC-99/RRASOR

The purpose of the validation study was to measure how well Static-99 predicted success and recidivism among PA DOC sex offenders. A valid risk assessment instrument will identify distinct groups of offenders with different likelihoods of re-offending. In other words, recidivism rate among the group of offenders identified as low risk will be significantly lower than recidivism rate among those identified as medium risk, which will be significantly lower than recidivism rate among high risk offenders.

A. Methodology

1. Sampling Method

The sample for the validation study included the 220 cases used in the reliability study, plus an additional 330 cases, for a total sample of 550 cases. These cases were randomly sampled from the total universe of all prisoners convicted of a sex offense who were released 1994-95 or 2000 as noted earlier. The sampling process resulted in 354 cases being assigned to 1994-95 release cohort and 194 for the 2000 releases. A comparison of the key demographic and legal status attributes of all sex offender releases and the sampled cases for these same time periods revealed no significant differences. Thus, the sample is considered to be representative of the population of sex offender parolees and the results can be generalized.

2. Data Collection and Analysis

Data extracted from PBPP files provided a profile of the sample's demographic and offense characteristics. In addition, Static-99/RRASOR data were available for all cases, which included responses to each item on the instrument, the total score and resultant assessed risk level. The offender's success on parole was tracked, using a 12-month follow-up period for the 2000 cohort and a six-year follow-up period for the 1994-95 cohort. Thus, we would expect that the 1994-95 cohort would have higher recidivism rates because of the extended time at-risk. For purposes of this study, the definition of recidivism included both arrests and returns to prison for a new conviction or for a technical violation as well as deaths due to criminality. Recidivism results are presented, both collapsed and separated across the two categories. These recidivism data were extracted from PBPP data files, which included the date of the recidivism event and the nature of the event. The PBPP forwarded the data to the ICJC, where it was converted an SPSS database for statistical analysis.

Descriptive (e.g., frequencies and cross-tabulations) and multivariate (e.g., logistic regression) analyses were conducted in order to produce tables that present demographic and offender characteristics, distribution of scores across items, and predictive attributes of the Static-99/RRASOR instruments.

B. Results

Table 5 provides a summary of the various outcomes for the sex offender releases for both cohorts. Overall, 48% of the cases were arrested or returned to prison during the follow-up period. However, there was a significant difference in the recidivism rates between the two cohort years—the 1994-95 cohort had a 55 percent recidivism rate and the 2000 cohort had a 36 percent recidivism rate. Most likely, this difference is due to the extended follow-up period for the 1994-95 cohort, versus the more limited 12 month follow-up period for the 2000 cohort. Of those who recidivated, approximately 70% (n=187) of these cases were returned for technical violations. Among the 52% of cases that are considered to be successful, most are reporting regularly or have had their sentence expire.

Table 5 Summary of Recidivism Measures Sex Offender Releases 1994-95 and 2000		
Outcome Measure	N	%
Total	550	100%
Successful	284	52%
Sentence Expired	110	20%
Reporting Regularly	165	30%
Death (Non-criminal)	9	2%
Unsuccessful (recidivism)	266	48%
Arrested or Recommitted for Technical Violation	187	34%
Arrested or Recommitted for New Crime	78	14%
Death (Criminally related)	1	0%

Tables 6 and 7 show the recidivism rates for each of the 10 scoring items for each release cohort. The recidivism rates are separated by technical violators and those convicted of a new crime. As shown in Table 6, 55 percent of the 1994-95 release cohort recidivated during the six-year follow-up period. The majority of these parolees returned to prison for technical violations. In general, recidivism rates were higher for offenders scoring in the “positive” direction on the items. More specifically, those with a prior sex offense were more likely to recidivate than those without a prior sex offense. Offenders with four or more prior sentencing dates, with convictions for non-contact sex offenses, with index non-sexual violent offenses, prior non-sexual violence convictions, unrelated victims,

stranger victims, and male victims were more likely to recidivate than their counterparts. Further, those younger than age 25 and those who did not have a history of cohabitation with a lover were more likely to recidivate than their older or co-habiting counterparts. The asterisked items in Table 6 identify the 4 RRASOR items. The recidivism rates for these 4 items (prior sex offenses, any unrelated victims, any male victims and young) are somewhat comparable when compared to the other 6 items. That is, nothing would be gained by using RRASOR as a separate instrument.

Table 7 depicts the same information as above, but for those prisoners released in 2000. The overall recidivism rate for this group is 36 percent, with the majority of offenders arrested or returned for technical violations (28.9%, versus 7.2% returned for a new conviction). Many of the same patterns identified above were found in the 2000 cohort. However, when looking only at the returns to prison for a new conviction, certain unexpected patterns emerged. For example, offenders with no prior sex offenses were returned to prison *more often* than offenders with prior sex offenses. Offenders without convictions for non-contact sex offenses were returned to prison *more often* than those with such convictions.

Despite these anomalies, the general trends conformed what would be expected intuitively. Looking again at the four RRASOR items separately in Table 7, these items predict recidivism at a slightly lower rate than the other 6 items. This could be due, however, to the fact that the short at-risk time of this cohort. As time goes on, the difference in recidivism percentages may begin to level off as illustrated by the 1994-1995 cohort.

One item that is problematic on for all of these cases is the age variable. The Static 99/RRASOR instrument uses a simple dichotomy to score this variable (age 25 or older versus those under age 25). For a variable to have any meaning, there must be variance in how the item is applied to prisoners eligible for release. Tables 6 and 7 illustrate that very few cases fall under the under age 25 designation and therefore has little impact on the assessment process. A re-analysis of the age category shows that the recidivism rate is fairly consistent among all age groups except ages 45 and older, where the recidivism rate drops dramatically. For this reason, it would be preferred to raise the age cut-off threshold from 25 to 45 years.

Table 8 presents the recidivism rates as defined in this study across the various risk levels. In general, recidivism rates across risk levels should conform to a linear pattern. In other words, low risk offenders should recidivate less often than medium risk offenders, who recidivate less often than high risk offenders. For the 1994-95 release cohort, this pattern generally holds, except that offenders assessed as high risk were less likely to recidivate than those assessed as medium risk. This pattern is also evident when comparing rates of return to prison for either a technical violation or a new conviction.

The lower rates for the high-risk cases may be related to problems in trying to score these parolees based on case files that inferior information on sex related offenses in a pre-Megan era. This problem was less present for the 2000 cohort, as the necessary information for scoring the instrument was easily located within the case files. For the 2000 release cohort, the expected linear pattern holds across all risk levels (Table 9).

C. Conclusions and Recommendations

Overall, Static-99 appears to be performing as expected. There appears to be a distinct difference in recidivism rates among most of the risk categories. The only recommendation we would make is that the age variable is changed as suggested above.

Table 6
Static-99 Items by Failure Variable - Years 1994-1995 Sex Offender Releases

Static-99 Item	Total		Recidivated	Technical	Convicted
	N	%	%	%	
	TOTAL	356	100.0%	55.1%	36.8%
Prior sex offenses (conv. or charges)?**					
None	252	70.8%	53.6%	37.7%	15.9%
1 or more (1 - 3)	104	29.2%	58.7%	34.6%	24.0%
Prior sentencing dates?					
3 or less (0)	254	71.3%	50.0%	36.2%	13.8%
4+ (1)	102	28.7%	67.6%	38.2%	29.4%
Any convictions for non-contact sex offender?					
No (0)	311	87.4%	54.3%	36.0%	18.3%
Yes (1)	45	12.6%	60.0%	42.2%	17.8%
Index non-sexual violence					
No (0)	230	64.6%	48.7%	32.2%	16.5%
Yes (1)	126	35.4%	66.7%	45.2%	21.4%
Prior non-sexual violence?					
No (0)	224	62.9%	48.7%	33.5%	15.2%
Yes (1)	132	37.1%	65.9%	42.4%	23.5%
Any unrelated victims? **					
No (0)	79	22.2%	44.3%	29.1%	15.2%
Yes (1)	277	77.8%	58.1%	39.0%	19.1%
Any stranger victims?					
No (0)	247	69.4%	50.6%	34.4%	16.2%
Yes (1)	109	30.6%	65.1%	42.2%	22.9%
Any male victims? **					
No (0)	318	89.3%	53.5%	35.5%	17.9%
Yes (1)	38	10.7%	68.4%	47.4%	21.1%
Young? **					
Aged 25 or older (0)	335	94.1%	54.6%	37.0%	17.6%
Aged 18-24.99 (1)	21	5.9%	61.9%	33.3%	28.6%
Single?					
Yes (1)	203	57.0%	50.7%	34.0%	16.7%
No (0)	153	43.0%	60.8%	40.5%	20.3%

Note: an ** is used to identify those items that comprise the RRASOR

Table 7
Static-99 Items by Failure Variable - Year 2000 Sex Offender Releases

Static-99 Item	Total		Recidivated	Technical	Convicted
	N	%	%	%	
	TOTAL	194	54.5%	36.1%	28.9%
Prior sex offenses (conv. or charges)?**					
None	161	83.0%	35.4%	27.3%	8.1%
1 or more (1 - 3)	33	17.0%	39.4%	36.4%	3.0%
Prior sentencing dates?					
3 or less (0)	149	76.8%	32.2%	26.2%	6.0%
4+ (1)	45	23.2%	48.9%	37.8%	11.1%
Any convictions for non-contact sex offender?					
No (0)	169	87.1%	34.9%	26.6%	8.3%
Yes (1)	25	12.9%	44.0%	44.0%	0.0%
Index non-sexual violence					
No (0)	140	72.2%	25.7%	20.7%	5.0%
Yes (1)	54	27.8%	63.0%	50.0%	13.0%
Prior non-sexual violence?					
No (0)	133	68.6%	33.8%	27.8%	6.0%
Yes (1)	61	31.4%	41.0%	31.1%	9.8%
Any unrelated victims? **					
No (0)	66	34.0%	25.8%	24.2%	1.5%
Yes (1)	128	66.0%	41.4%	31.3%	10.2%
Any stranger victims?					
No (0)	150	77.3%	30.7%	26.7%	4.0%
Yes (1)	44	22.7%	54.5%	36.4%	18.2%
Any male victims? **					
No (0)	170	87.6%	37.6%	30.0%	7.6%
Yes (1)	24	12.4%	25.0%	20.8%	4.2%
Young? **					
Aged 25 or older (0)	191	98.5%	36.1%	28.8%	7.3%
Aged 18-24.99 (1)	3	1.5%	33.3%	33.3%	0.0%
Single?					
Yes (1)	129	66.5%	31.8%	27.1%	4.7%
No (0)	65	33.5%	44.6%	32.3%	12.3%

Note: an ** is used to identify those items that comprise the RRASOR

Table 8
Current Static-99 Score by Failure Variable - Years 1994-1995 Sex Offender Releases

Point Distribution		Total		Recidivated	Technical	Convicted
		N	%	%	%	%
Low	0	30	8.4%	26.7%	13.3%	13.3%
	1	55	15.4%	36.4%	32.7%	3.6%
	Total Low Risk	85	23.9%	32.9%	25.9%	7.1%
Med-Low	2	47	13.2%	53.2%	42.6%	10.6%
	3	66	18.5%	56.1%	33.3%	22.7%
	Total Med-Low Risk	113	31.7%	54.9%	37.2%	17.7%
Med-High	4	62	17.4%	69.4%	43.5%	25.8%
	5	37	10.4%	73.0%	48.6%	24.3%
	Total Med-High Risk	99	27.8%	70.7%	45.5%	25.3%
High	6	34	9.6%	52.9%	32.4%	20.6%
	7	15	4.2%	66.7%	40.0%	26.7%
	8	7	2.0%	71.4%	42.9%	28.6%
	9	2	0.6%	100.0%	50.0%	50.0%
	10	1	0.3%	100.0%	100.0%	0.0%
	Total High Risk	59	16.6%	61.0%	37.3%	23.7%

Table 9
Current Static-99 Score by Failure Variable - Year 2000 Sex Offender Releases

Point Distribution		Total		Recidivated	Technical	Convicted
		N	%	%	%	%
Low	0	27	13.9%	18.5%	18.5%	0.0%
	1	40	20.6%	30.0%	22.5%	7.5%
	Total Low Risk	67	34.5%	25.4%	20.9%	4.5%
Med-Low	2	37	19.1%	24.3%	18.9%	5.4%
	3	31	16.0%	48.4%	41.9%	6.5%
	Total Med-Low Risk	68	35.1%	35.3%	29.4%	5.9%
Med-High	4	22	11.3%	40.9%	31.8%	9.1%
	5	25	12.9%	48.0%	40.0%	8.0%
	Total Med-High Risk	47	24.2%	44.7%	36.2%	8.5%
High	6	7	3.6%	71.4%	28.6%	42.9%
	7	3	1.5%	66.7%	66.7%	0.0%
	8	2	1.0%	50.0%	50.0%	0.0%
	9	0	0.0%	0.0%	0.0%	0.0%
	10	0	0.0%	0.0%	0.0%	0.0%
	Total High Risk	12	6.2%	66.7%	41.7%	25.0%