By

R. Karl Hanson
Department of the Solicitor General of Canada, Ottawa

David Thornton
Her Majesty's Prison Service, London

Static 99: Improving Actuarial
Risk Assessments for Sex Offenders

1999-02

AUTHOR NOTE

Correspondence concerning this article can be addressed to either author.

R. Karl Hanson
Corrections Research
Department of the Solicitor General of Canada
340 Laurier Ave., West
Ottawa, Ontario
Canada
K1A 0P8

Telephone: (613) 991 2840 Fax: (613) 990 8295
email hansonk@sgc.gc.ca

David Thornton
Offender Behaviour Programmes Unit
Room 701
HM Prison Service
Abell House
John Islip Street
London SW1P 4LH

Telephone: (171) 217 5370 Fax: (171) 217 5871
email DavidThornton1@compuserve.com

Abstract

The study compared the predictive accuracy of three sex offender risk assessment measures: the RRASOR (Hanson, 1997), Thornton's SACJ-Min (Grubin, 1998), and a new scale, Static-99, created by combining the items from the RRASOR and SACJ-Min. Predictive accuracy was tested using four diverse data sets drawn from Canada and the UK (total n = 1,301). The RRASOR and the SACJ-Min showed roughly equivalent predictive accuracy and the combination of the two scales was more accurate than either original scale. Static-99 showed moderate predictive accuracy for both sexual recidivism (r = .33, ROC area = .71) and violent (including sexual) recidivism (r = .32, ROC area = .69). The variation in the predictive accuracy of Static-99 across the four samples was no more than would be expected by chance.

STATIC 99: IMPROVING ACTUARIAL
RISK ASSESSMENTS FOR SEX OFFENDERS

The management of sex offenders within the criminal justice system can be substantially influenced by the offender's perceived risk for recidivism. Those sex offenders deemed high risk may be subject to substantial restrictions, such as post-sentence detention, indeterminate sentences, and long-term community supervision. Conversely, sex offenders deemed to be low risk may be placed on probation and, if incarcerated, be considered for early release.

Although many decisions require risk assessments, the procedures used for making such assessments often have limited validity. In general, the average predictive accuracy of professional judgement to predict sex offence recidivism is only slightly better than chance (average r = .10, Hanson & Bussière, 1998). Some have even argued that the accuracy of prediction is sufficiently low that it threatens the very basis of risk-based legal sanctions for sex offenders (Janus & Meehl, 1997).

Recent research, however, has the potential of substantially improving the accuracy of recidivism risk assessments for sex offenders. Hanson and Bussière's (1998) meta-analytic review identified a number of risk factors that were reliably associated with sex offence recidivism. Most of these factors were static, historical variables related to sexual deviance (e.g., prior sex offences, stranger victims) and general criminality (e.g., prior non-sex offences, antisocial personality disorder). Several different actuarial risk instruments have also been developed to predict recidivism among sexual offenders (e.g., Sex Offender Risk Appraisal Guide [SORAG], Quinsey, Harris, Rice & Cormier, 1998; Minnesota Sex Offender Screening Tool – Revised [MnSOST-R], Epperson, Kaul & Hesselton, 1998); Rapid Risk Assessment for Sex Offence Recidivism [RRASOR], Hanson, 1997; Thornton's Structured Anchored Clinical Judgement [SACJ], Grubin, 1998). These actuarial scales not only specify the items to consider, but also provide explicit direction as to the relative importance of each item. The items in the scales are similar, although the scales vary as to the relative weight accorded to the general factors of sexual deviance versus antisociality.

The SORAG (Quinsey et al., 1998) is a variation of the Violence Risk Appraisal Guide (VRAG; Quinsey et al., 1998) for sexual offenders. Like the VRAG, the SORAG was designed to assess any violent recidivism, not just sexual recidivism. It contains 15 items addressing early childhood behaviour problems, alcohol problems, sexual and nonsexual criminal history, age, marital status, and personality disorders (with a large weight on psychopathy). The MnSOST-R was developed to predict sexual recidivism among rapists and extrafamilial child molesters. The MnSOST-R includes 16 items addressing sexual and non-sexual criminal history, the victims' age and relationship to the offender, substance abuse, unstable employment, age, and treatment history (Epperson et al., 1998). Both the RRASOR (Hanson, 1997) and SACJ (Grubin, 1998) were intended to be relatively brief screening instruments for predicting sexual offence recidivism.

The purpose of the present study was to compare the predictive accuracy of two of these actuarial schemes: the RRASOR (Hanson, 1997) and the SACJ (see Grubin, 1998). Although rarely used in North American, the SACJ is routinely used in Her Majesty's Prison Service

(England and Wales) and in many police departments in the UK. The SACJ contains items related to sexual deviance, but also places considerable weight on non-sexual criminal history. The RRASOR, in contrast, almost exclusively targets factors related to sexual deviance. The RRASOR is widely used in Canada and the U.S., being the most common risk assessment tool used in post-sentence detention procedures (Doren, 1999). Given the different emphasis of the RRASOR and SACJ, one goal of the current study was to examined whether a simple combination of these two scales could improve upon the predictive accuracy of either original scale.

Rapid Risk Assessment for Sex Offence Recidivism (RRASOR; Hanson, 1997)

The aim of the RRASOR was to predict sex offence recidivism using a small number of easily scored variables. The initial pool of seven items were those that correlated at least .11 with sex offence recidivism in Hanson and Bussière's (1998) meta-analysis and were commonly recorded: prior sex offences, any prior non-sex offences, any male victims, any stranger victims, any unrelated victims, never married, and age less than 25 years. In order to identify the most efficient combination of these items, the correlations between these predictor variables were calculated in seven different data sets (total sample of 2,592), and then averaged using standard meta-analytic techniques (Hedges & Olkin, 1985). Following a suggestion by Becker (1996), the averaged correlation matrix was then subjected to step-wise regression to identify the best predictor variables.

Of the original seven variables, four substantially contributed to the regression equation (beta greater than .09): prior sex offences, any unrelated victims, any male victims and age less than 25 (see Table I). The scale resulting from the simple combination of these four variables was then tested on an entirely new sample (HM Prison). Overall, the scale showed comparable predictive accuracy in both the development and validation samples (average r = .27; average ROC area = .71).

Structured Anchored Clinical Judgement (SACJ; Grubin, 1998).

The SACJ aims to predict sexual and violent recidivism using a stage approach, with each stage incorporating different types of information. The first stage considers the offender's official convictions: specifically, any current sex offences, any prior sex offences, any current non-sexual violent offences, any prior non-sexual violent offences, and four or more prior sentencing occasions (see Table 1). If offenders have four or more of the initial factors, they are automatically considered high risk. If two or three factors are present, offenders are considered medium risk, and zero or one factors indicate low initial risk.

Table 1

Items in the RRASOR, SACJ-Min, and Static-99

Type of risk factor	RRASOR	SACJ-Min	Static-99
Sexual deviance	male victims	male victims	male victims
		never married	never married
		non-contact sex offences	non-contact sex offences
Range of potential victims	unrelated victims		unrelated victims
		stranger victims	stranger victims
Persistence	prior sex offences (3 points)		prior sex offences (3 points)
		current sex offence	
		prior sex offence	
Antisociality		current non-sexual violence	current non-sexual violence
		prior non-sexual violence	prior non-sexual violence
		4+ sentencing dates	4+ sentencing dates
Age	18 - 24.99 years		18 - 24.99 years

The second step considers a number of potentially aggravating factors, such as lack of prior relationship to victim. If two or more of these factors are present, then the offenders' initial risk level is increased one category. The eight potentially aggravating factors are divided into two sets. Set A includes any stranger victims, any male victims, never married, and convictions for non-contact sex offences (e.g., exhibitionism, obscene phone calls). Set B includes items that are somewhat more difficult to assess such as substance abuse, placement in residential care as a child, deviant sexual arousal, and psychopathy. The SACJ was designed to be used even when there is missing data. The Step 1 and Step 2 - Set A items are considered the minimum required for a valid assessment, and using these items results in a reduced scale called SACJ-Min.

The final step of the SACJ (Step 3) considers information that is unlikely to be obtained except for sex offenders who enter treatment programs (e.g., treatment drop-out, improvement on dynamic risk factors). Since only the SACJ-Min has been subject to cross-validation, the final step of the SACJ will not be considered further in this report.

The SACJ was developed through exploratory analyses on several UK data sets. The SACJ-Min was then validated on an entirely new sample of approximately 500 sex offenders released from Her Majesty's Prison Service in 1979 (16 year follow-up on the complete cohort). This HM Prison sample included the 303 offenders originally used to validate the RRASOR. In the validation sample, the SACJ-Min correlated .34 with sex offence recidivism and .30 with any sexual or violent recidivism (Thornton, Personal communication, February 10, 1999). The SACJ-Min has yet to be tested on samples from outside the UK.

Static-99

Preliminary analyses suggested that RRASOR and the SACJ-Min were assessing related, but not identical constructs. Both scales contributed unique variance to regression equations when their total scores were used to predict sexual recidivism. Consequently, it was possible that a combination of the two scales may predict better than either original scale. A new scale was created by adding together the items from the RRASOR and SACJ-Min. The scale is called Static-99 to indicate that it includes only static factors and that the current version is this year's version of a work in progress. The complete list of items is listed in Table 1 and the scoring criteria are given in Appendix I.

Importance of replication

It is important that risk scales developed on one sample be tested on at least one independent sample. Without replication, the relationships found in the development sample may be related to idiosyncratic features of that sample. Evaluators applying a risk scale to new settings would have increased confidence if the scale had already been demonstrated to show adequate predictive accuracy in a variety of settings.

Replications, however, are more often advocated than conducted. The observed sex offence recidivism base rate is sufficiently low that many years are required before new studies yield meaningful results. Researchers eager for new results have the option of using existing data bases, but data bases created for one purpose may poorly fit other needs. Apart from the

obvious problem of missing variables, different data sets often have subtle variations in the definitions of the variables. For example, recidivism may be defined by charges versus convictions, or the relationship to victims may be based on officially recorded offences versus for all known offences.

When a risk scale shows significant variability across samples, the differences may be due to variation in scoring procedures, or the scale may have differential validity in different samples. On the other hand, if similar results are found across samples (despite minor differences in coding rules), then a scale would appear robust.

METHOD

<u>Samples</u>

The first three samples were, with minor modifications, the same samples used in the development of the RRASOR (see Table 2). The results reported below are not identical to those reported in Hanson (1997) due to minor recoding of some variables (correcting coding errors, replacing missing data). The fourth sample (HM prison) was not used in the development of either the RRASOR or SACJ, but a subsample of the HM Prison offenders were used as the validation sample for both these risk scales. The HM Prison sample has the important feature of being an unbiased cohort of all the sex offenders released in the target year (1979). In contrast, the other samples primarily comprised sex offenders referred to assessment and/or treatment at particular institutions.

<u>Institut Philippe Pinel (Montreal)</u>. (Proulx, Pellerin, McKibben, Aubut & Ouimet, 1995; see also Proulx, Pellerin, McKibben, Aubut & Ouimet, 1997; Pellerin et al., 1996). This study focused on sexual offenders treated at a maximum security psychiatric facility between 1978 and 1993. The Institut Philippe Pinel provides long term (1-3 years) treatment for sexual offenders referred from both the mental health and correctional systems. Information concerning predictor variables was drawn from their clinical files and recidivism information from RCMP records collected in 1994.

Information was available on all the predictor variables except stranger victims and non-contact sexual offences. As well, it was impossible to separate index and prior non-sexual violence since only the total number of charges for non-sexual violence were recorded. Similarly, the variable marking the total number of sex offence charges included index offences. To estimate the number of prior sex offence convictions, the number of victims for the index offence was subtracted from the total number of charges.

<u>Millbrook Recidivism Study</u> (Hanson, Steffy & Gauthier, 1993b; see also Hanson, Scott & Steffy, 1995; Hanson, Steffy & Gauthier, 1992; Hanson, Steffy & Gauthier, 1993a). This study collected long-term recidivism information (15-30 years) for child molesters released between 1958 and 1974 from Millbrook Correctional Centre, a maximum security provincial correctional facility located in Ontario, Canada. About half of the sample went through a brief treatment program. For the treatment sample, the information concerning the predictors was collected from their clinical files, whereas for the remainder of the sample, the information was extracted from their correctional files. Recidivism information was coded from national records maintained by the Royal Canadian Mounted Police (RCMP).

Information was available on all the relevant predictor variables, except for convictions for non-contact sex offences (missing for all cases). Information concerning stranger victims was available for the treatment sample only (n = 99). As well, the total number of prior convictions was used instead of the total number of prior sentencing dates.

<u>Oak Ridge Division of the Penetanguishene Mental Health Centre</u>. (Rice & Harris, 1996; see also Quinsey, Rice & Harris, 1995; Rice & Harris, 1997; Rice, Harris & Quinsey, 1990; Rice,

Quinsey & Harris, 1991). The Oak Ridge study followed sexual offenders referred between 1972 and 1993 for treatment and/or assessment to a maximum security mental health centre located in Ontario, Canada. The majority of the referrals came from the mental health systems or the courts (e.g., pretrial fitness examinations), with a minority of cases coming from provincial or federal corrections. Follow-up information was based on RCMP records as well as mental health records (i.e., new admissions for sexual offenses, whether or not new charges were laid).

Information was available for all the predictor variables with the following exceptions. Convictions for non-contact sex offence was not available for all cases. Relationship to victim was only available for the most serious offence. The data set counted any male child victims rather than any male victims. The number of prior convictions was used instead of the number of prior sentencing dates. Finally, only the most serious index offence was recorded in the data set. Consequently, index convictions for non-sexual violence that was considered less serious than the index sex offence would not have been recorded.

Her Majesty's Prison Service (UK). (Thornton, 1997). The study provided a 16 year follow-up of 563 sexual offenders released from Her Majesty's Prison Service (England and Wales) in 1979. Recidivism information was based on Home Office records collected in 1995. Very few of the offenders in this sample would have received specialised sexual offender treatment.

Information was available for all the relevant predictor variables. Previous sex offences, however, was coded based on the offenders' previous sentencing occasions rather than the number of convictions or charges.

Table 2

	S A M P L E			
	Institut Philippe Pinel	**Millbrook**	**Oak Ridge**	**HM Prison England and Wales**
Setting	secure psychiatric	provincial prison	secure psychiatric	all prisoners released in 1979
Minimum Sample Size	344	191	142	531
Age at Release (SD)	36.2 (10.9)	33.1 (9.9)	30.4 (9.5)	34.4 (12.7)
% Child Molesters	70.4	100.0	49.3	60.7
Prior Offences				
Sexual (%)	50.5	41.9	31.8	34.0
Any (%)	58.1	72.0	67.7	74.9
Average Years of Follow-up	4	23	10	16
Recidivism Criteria	convictions	convictions	charges/ readmissions	convictions
Recidivism rates				
Sexual Only (%)	15.4	35.1	35.1	25.0
Any violent (%)	21.5	44.0	57.6	37.4

ANALYSIS

Measure of predictive accuracy

The area under the Receiver Operating Characteristic (ROC) curve was used as the primary measure of predictive accuracy (Hanley & McNeil, 1982; Mossman, 1994; Rice & Harris, 1995). The ROC curve plots the hits (accurately identified recidivists) and false alarms at each level of the risk scale. The area under the ROC curve can range from .50 to 1.0, with 1.0 indicating perfect prediction (no overlap between recidivists and non-recidivists) and .50 indicating prediction no better than chance. In general, the ROC area can be interpreted as the probability that a randomly selected recidivist would have a more deviant score than a randomly selected nonrecidivist. The ROC area has advantages over other commonly-used measures of predictive accuracy (e.g., percent agreement, correlation coefficients, RIOC) since it is not constrained by base rates or selection ratios (see Swets, 1986).

The correlation coefficient, r, is also presented to facilitate comparison with the results of other studies. For example, the average correlation between prior sex offences and sex offence recidivism is .19 (95% confidence interval .17 to .21; Hanson & Bussière, 1998). To have utility in predicting long-term recidivism, risk scales need to improve upon this minimum standard.

Comparing results

Standard meta-analytic procedures were used to compare results across studies (Hedges & Olkin, 1985; Hedges, 1994; McClish, 1992). Variability across studies was indexed by the Q statistic: $Q = \sum w_i (A_i - A.)^2$, where A_i is the ROC area for each sample, w_i is the weight for each sample (inverse of its variance – SE^2), and A. is the weighted grand mean ($\sum w_i A_i / \sum w_i$). The Q statistic is distributed as χ^2 with degrees of freedom equal to k - 1, where k is the number of groups. The predictive accuracy of the risk scales were compared using the test of correlated ROC areas described by Hanley and McNeil (1983): $Z = (A_1 - A_2)/(SE_1^2 + SE_2^2 - 2rSE_1SE_2)^{1/2}$. The ROC statistics were computed using ROCKIT Version 0.9.1 (Metz, 1998).

Estimating recidivism rates

Applied risk assessments are often concerned about whether offenders have a specific probability of recidivism (e.g., greater than 50%). Since recidivism rates are highly influenced by the length of the follow-up period, recidivism probabilities were estimated using survival analysis (Allison, 1984; Soothill & Gibbens, 1978). Survival analysis calculates the probability of recidivating for each time period given that the offender has not yet reoffended. Once offenders recidivate, they are removed from the analysis of subsequent time periods. Survival analysis has the advantage of being able to estimate year by year recidivism rates even when the follow-up periods vary across offenders. Readers should be aware, however, that the estimates for the longest follow-up periods can be unstable if there are few offenders remaining in the later years.

RESULTS

As can be seen in Table 3, the predictive accuracy of the scales was relatively consistent across the samples. For both the RRASOR and Static-99, the amount of variability was no greater than would be expected by chance (all p > .30). The SACJ-Min, however, showed significant variability in the prediction of sexual recidivism (Q = 7.89, df = 3, p < .05). The SACJ-Min predicted sex offence recidivism best in HM Prison sample (A = .74) and worst in the Millbrook sample (A = .61).

Table 3

Predictive accuracy of RRASOR, SACJ-Min, and Static-99 across samples (ROC areas)

	Pinel	Millbrook	Oak Ridge	HM Prison 1979	Average		
					A.	Q	Sample Size
Sexual Recidivism							
RRASOR	.71	.66	.62	.71	.68	3.56	1,225
SACJ-Min	.66	.61	.63	.74	.69	7.89*	1,301
Static-99	.73	.65	.67	.72	.70	3.42	1,228
Any Violent Recidivism							
RRASOR	.65	.67	.60	.65	.65	1.17	1,228
SACJ-Min	.65	.65	.67	.69	.67	2.24	1,304
Static-99	.71	.71	.69	.69	.69	1.52	1,231

*p < .05.

The samples were combined to directly test the relative predictive accuracy of the RRASOR, SACJ-Min and Static-99 (see Table 4). Only subjects who had complete data on all three risk scales were used in the combined sample (total n = 1,208). The average values of the scales in the combined samples were as follows: RRASOR mean = 1.77, SD = 1.29; SACJ-Min, mean = 2.02, SD = .76; Static-99 mean = 3.15, SD = 1.97. The comparison of predictive accuracy of the scales used the test for correlated ROC areas described by Hanley and McNeil (1983).

Table 4

Relative predictive accuracy of the RRASOR, SACJ-Min and Static-99.

	Combined Sample (n = 1,208)				Rapists (n = 363)	Child molesters (n = 799)
	ROC Area	95% C.I.	r	95% C.I.	ROC area	ROC area
Sexual recidivism						
RRASOR	.68	.65-.72	.28	.23-.33	.68	.69
SACJ-Min	.67	.63-.71	.23	.18-.28	.69	.68
Static-99	.71	.68-.74	.33	.28-.38	.71	.72
Any violent Recidivism						
RRASOR	.64	.60-.67	.22	.16-.27	.64	.66
SACJ-Min	.64	.61-.68	.22	.16-.27	.62	.66
Static-99	.69	.66-.72	.32	.27-.37	.69	.71

For the prediction of sex offence recidivism, Static-99 (A = .71) was more accurate than the RRASOR (A = .68, Z = 2.38, p < .05) or the SACJ-Min (A = .67, Z = 2.84, p < .01). The RRASOR and SACJ-Min predicted sex offence recidivism with similar levels of accuracy (Z = .72, p > .40). For the prediction of any violent recidivism (including sexual), Static-99 (A = .69) was more accurate than either the RRASOR (A = .64, Z = 5.37, p < .001) or SACJ-Min (A = .64, Z = 3.84, p < .001). The RRASOR and SACJ-Min did not differ in the accuracy with which they predicted violent recidivism (Z = .35, p > .70).

In order to test the generalisability of the scales across subgroups of sex offenders, the offenders was divided into those who victimised adult females (rapists, n = 363) and those who victimised children (child molesters, n = 799). The comparison of predictive accuracy across these groups used the test of uncorrelated ROC areas described by McClish (1992). All the scales showed similar predictive accuracy for both rapists and child molesters (all Z < 1, all p > .30).

As can be seen from Figure 1 and Figure 2, the recidivism rates were very similar in the Pinel, HM Prison and Millbrook samples (for sexual recidivism, Survival χ^2 = 1.62, df = 2, p > .40; for violent recidivism, Survival χ^2 = .65, df = 2, p > .70). Survival dates were not available for the Oak Ridge sample. Given the similarity in the samples, the three data sets (Pinel, HM Prison, Millbrook) were combined for the purpose of creating estimated recidivism rates.
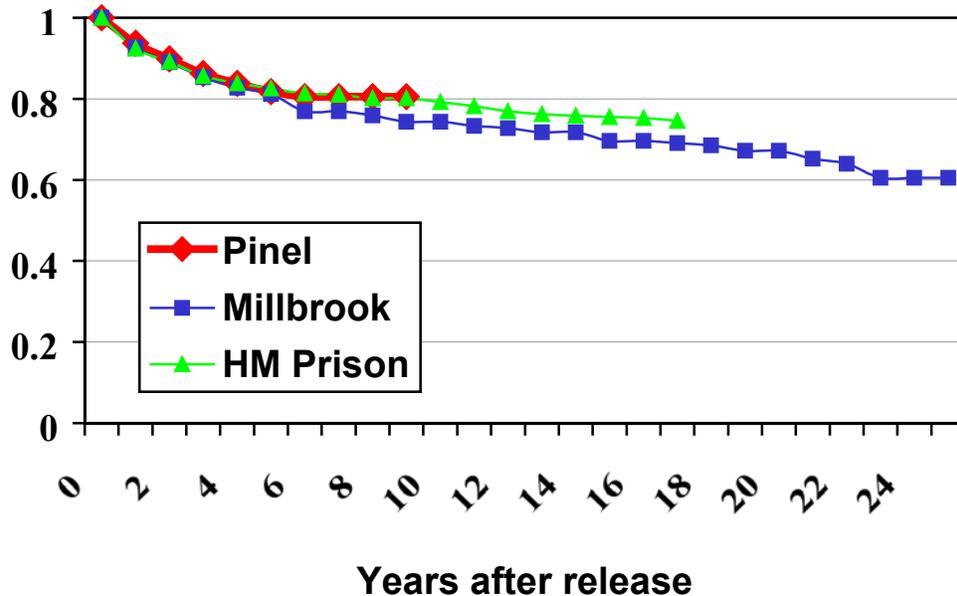

Figure 1. Sex offence recidivism rates (survival curves) for offenders released from three institutions.
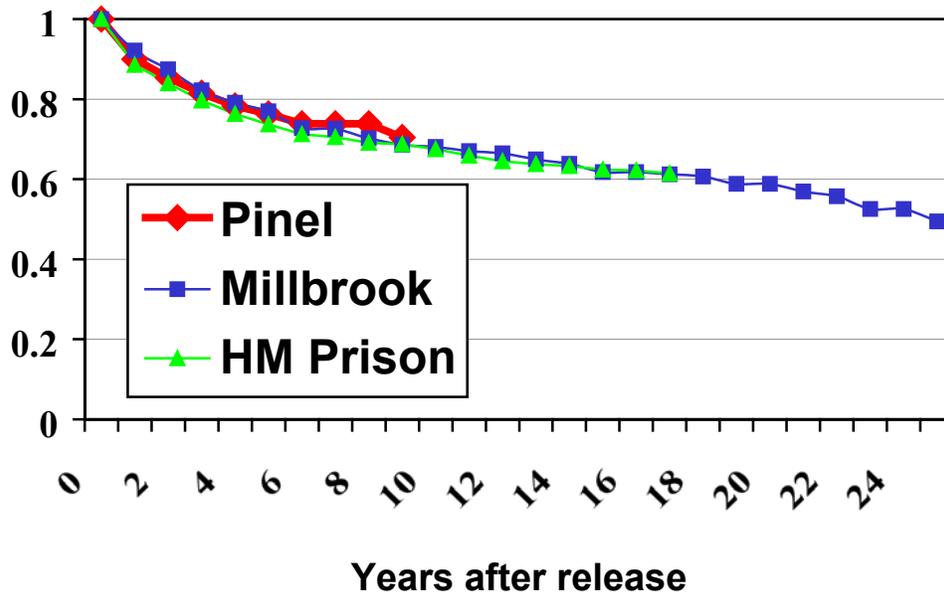
Figure 2. Violent recidivism rates (survival curves) for offenders released from three
institutions.

The relationship between Static-99 scores and sexual recidivism is presented in Figure 3.
The Static-99 scores were categorised as Low (0, 1; n = 257), medium-low (2, 3; n = 410),
medium-high (4, 5; n = 290) and high (6 plus; n = 129). To minimise the influence of isolated,
late recidivism events, the survival curves ended when there were fewer than 15 offenders
exposes to risk for a particular year. The observed 5, 10 and 15 year recidivism rates are
presented in Table 5. The rates up to 15 years should be reasonably reliable since all the
offenders in the HM Prison and Millbrook samples were followed for at least 15 years.

Figure 3. The relationship of Static-99 scores to sexual recidivism.

Figure 4. The relationship of Static-99 socres to violent recidivism.

Table 5

Recidivism rates for Static-99 risk levels.
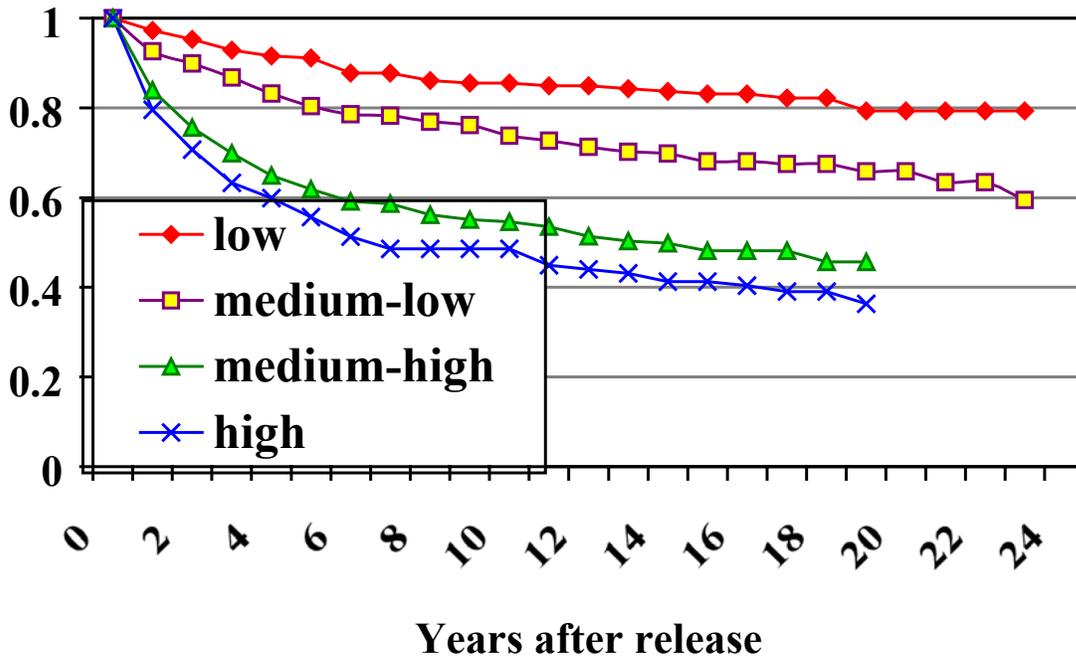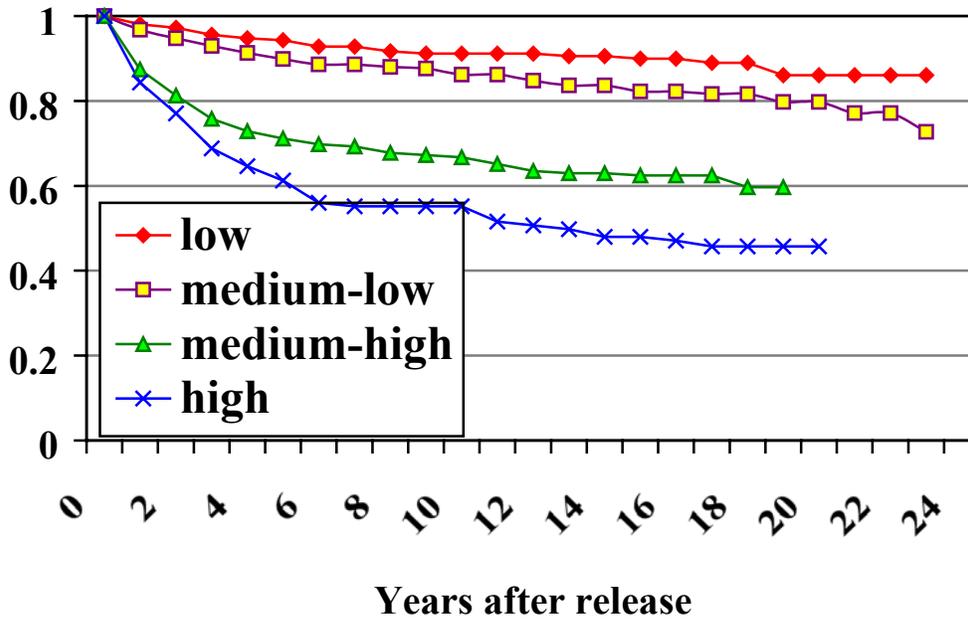
Static-99 score	Sample size	Sexual recidivism			Violent recidivism		
		5 years	10 years	15 years	5 years	10 years	15 years
0	107 (10%)	.05	.11	.13	.06	.12	.15
1	150 (14%)	.06	.07	.07	.11	.17	.18
2	204 (19%)	.09	.13	.16	.17	.25	.30
3	206 (19%)	.12	.14	.19	.22	.27	.34
4	190 (18%)	.26	.31	.36	.36	.44	.52
5	100 (9%)	.33	.38	.40	.42	.48	.52
6 +	129 (12%)	.39	.45	.52	.44	.51	.59
Average 3.2	1086 (100%)	.18	.22	.26	.25	.32	.37

Static-99 identified a substantial subsample (approximately 12%) of offenders whose long-term risk for sexual recidivism was greater than 50%. The recidivism rates for the minimum entrant into the high risk category (score of '6') was 37%, 44% and 51% after 5, 10 and 15 years post release. Most of the offenders, however, were in the lower risk categories, with long-term recidivism risk of 10% to 20%.

As can be seen in Figure 4, offenders with high scores on Static-99 were also at substantial risk for any violent recidivism (approximately 60% violent recidivism rate over 15 years). The violent recidivism rate (including sexual) for the minimum entrant into the high risk category (score of '6') was 46%, 53% and 60% over 5, 10, 15 years, respectively. The violent recidivism rate of Static-99's Low risk category (0, 1) was 17% after 15 years.

DISCUSSION

The study compared the predictive accuracy of three sex offender risk assessment measures (the RRASOR, the SACJ-Min, and a combined scale, Static-99) across four data sets. The RRASOR and the SACJ-Min showed roughly equivalent predictive accuracy and the combination of the two scales was more accurate than either original scale. The incremental improvement of Static-99, however, was relatively small. Static-99 showed moderate predictive accuracy for both sexual recidivism (r = .33, ROC area = .71) and violent (including sexual) recidivism (r = .32, ROC area = .69). The variation in the predictive accuracy of Static-99 across the four samples was no more than would be expected by chance.

If a risk scale is to be used in applied contexts, then it is important to considered whether the degree of predictive accuracy is sufficient to inform rather than mislead. Critics could suggest, for example, that a correlation in the .30 range is insufficient for decision-making since it only accounts for 10% of the variance. Even if such an argument was correct (and many argue that it is not – see Ozer, 1985), most decision-makers are not particularly concerned about "percent of variance accounted for". Instead, applied risk decisions typically hinge on whether offenders surpass a specified probability of recidivism (e.g., more than 50%).

Estimating absolute recidivism rates is a difficult task since many sex offences go undetected (e.g., Bonta & Hanson, 1994). Observed recidivism rates (especially with short follow-up periods) are likely to substantially underestimate the actual recidivism rates. Nevertheless, Static-99 identified a substantial subsample of offenders (approximately 12%) whose observed sex offence recidivism rate was greater than 50%. At the other end, the scale identified another subsample whose observed recidivism rates was only 10% after 15 years. Differences of this magnitude should be of interest to many applied decision-makers.

The similarity in the observed recidivism rates across the samples allows some confidence in conviction rate estimates provided by Static-99. The degree of similarity was remarkable considering that the studies were drawn from different countries, different language groups, different settings (i.e., prison, secure hospital), and different decades. All the studies for which survival data was available used official conviction as the outcome criteria. On the other hand, the Oak Ridge sample had a higher recidivism rate than the other three samples. Thirty-five percent of the Oak Ridge sample recidivated with a sex offence recidivism rate within 10 years, whereas only 25% of the HM Prison Service recidivated after a longer follow-up period (16 years). The Oak Ridge recidivism rates were relatively high since they used a broad recidivism criteria (arrests, re-admissions) and they may have included particularly high risk offenders. In support of the later hypothesis, Scheffé's post hoc tests found that the mean score on Static-99 was higher in the Oak Ridge sample (mean = 4.1) than in the other three samples (mean = 3.0). Whether recidivism rate differences would remain after controlling for pre-existing risk levels could not be determined with the available data.

Another approach to judging a measure's predictive accuracy is to compare it to the available alternatives. For the prediction of sex offense recidivism, Static-99 is clearly more accurate (r = .33) than unstructured clinical judgement (average r = .10; Hanson & Bussière,

1998).   The Violence Risk Appraisal Guide (VRAG), one of best established risk assessment instruments, correlated only .20 with sex offence recidivism in a cross-replication (Rice & Harris, 1997). Quinsey et al. (1998) have proposed a revision of the VRAG for sexual offenders, entitled the Sex Offender Risk Appraisal Guide (SORAG). Although the SORAG is reported to be a good predictor of violent recidivism, its relationship to sexual recidivism is relatively weak (ROC area of .62 compared to .67 for Static-99 in the same Oak Ridge data set). The MnSOST-R appears to predict sex offence recidivism (r = .45) somewhat better than Static-99, but the Min-SOST has yet to be fully cross-validated (Epperson et al., 1998).

Although Static-99 was designed to predict sex offence recidivism, it also showed reasonable accuracy in the prediction of any violent recidivism among sex offenders (r = .32, ROC area = .69). In comparison, a recent meta-analysis found the average correlation between Hare's Psychopathy Checklist-Revised (Hare, 1991) and violent recidivism was .27 (n = 1,374; Hemphill, Hare & Wong, 1998). Static-99, however, may not be the instrument of choice when the goal is predicting any violent recidivism. The VRAG, for one, predicts any violent recidivism substantially better than the Static-99 (r = .47, ROC area = .77, in a cross-replication sample of 159 sex offenders, Rice & Harris, 1997). Nevertheless, Static-99 may be useful in settings that lack the time, resources and/or information required to complete the VRAG.

The combination of the RRASOR and SACJ-Min was called Static-99 to indicate that it includes only static variables, and that it is this year's version of a work in progress. It is likely that actuarial risk scales can improve upon Static-99 by including dynamic (changeable) risk factors as well as additional static variables. The variables in Table 1 are grouped according to five dimensions that are plausibly related to the risk of sexual offence recidivism: sexual deviance, range of available victims, persistence (lack of deterrence or "habit strength"), antisociality, and age (young). The variables chosen to mark these dimensions were those conveniently available in the existing data sets. Deliberate efforts to create variables targeting these risk dimensions has the promise of substantially improving the prediction of sex offence recidivism. Additional variables could include, for example, repetitive victim choice (same age and sex) as a marker for sexual deviance (see Freund & Watson, 1991) or early onset of sex offending as a marker of "persistence".

The inclusion of dynamic factors would likely increase the scale's predictive accuracy (Hanson & Harris, 1998, in press). Among non-sexual criminals, dynamic variables predict recidivism as well or better than static variables (Gendreau, Little & Goggin, 1996).   The research on dynamic factors related to sexual offending is not well developed, but some plausible dynamic risk factors include intimacy deficits (Saidman, Marshall, Hudson & Robertson, 1994), sexualisation of negative affect (Cortoni, 1998), attitudes tolerant of sexual assault (Hanson & Harris, 1998), emotional identification with children (Wilson, 1999), treatment failure, and non-cooperation with supervision (Hanson & Harris, 1998).

Use of Static-99 in sex offender risk assessments.

The Static-99 is intended to be a measure of long-term risk potential. Given its lack of dynamic factors, it cannot be used to select treatment targets, measure change, evaluated whether offenders have benefited from treatment, or predict when (or under what circumstances) sex offenders are likely to recidivate.

There are several different ways in which empirically derived risk scales can be used in clinical assessments. Quinsey et al. (1998) have argued for a pure actuarial approach: risk predictions are those provided by the actuarial scale with no allowances for other factors. They argue that clinical judgement is so much inferior to actuarial methods that any consideration of clinical judgement simply dilutes predictive accuracy.

Their position is plausible and is likely true in many situations. However, actuarial risk scales are accurate to the extent that they consider all relevant risk factors. Static-99 does not claim to be comprehensive, for it neglects whole categories of potentially relevant variables (e.g., dynamic factors). As well, prudent evaluators would want to consider whether there are special features of individual cases that limit the applicability of actuarial risk scales (e.g., a debilitating disease or stated intentions to reoffend).

As research progresses, variables external to the actuarial scheme will either be shown to improve risk predictions (and be incorporated into scales) or be shown to add no new information and be dismissed. Until the desired empirical information is available, evaluators wishing to consider external variables need to carefully articulate the rationale for including each variable. One plausible approach is to begin with the risk predictions provided by the actuarial scale, and adjust these predictions (up or down) based on empirically validated risk factors that were not considered in the development of the original actuarial scale. In most cases, the optimal adjustment would be expected to be minor or none at all.

The Structured Risk Assessment (SRA) framework developed by David Thornton is one example of a structured approach to combining actuarial risk scales with other empirically based risk factors. The current version of SRA uses Static 99 as the first step in risk assessment. The second step uses the offenders' functioning on dynamic risk factors to revise this initial classification. Medium risk cases are re-classified as high risk if their functioning is psychologically similar to high risk offenders, and it is reclassified down to lower risk if their functioning is psychologically similar to low risk offenders. The third step uses information devised from response to treatment. The fourth step considers the offenders' typical offence pattern in conjunction with situational risk factors. This kind of system reflects the complexity of the real situations in which risk assessment takes place. At each stage the system is empirically based, becoming actuarial where practical and elsewhere using lesser, although still credible, forms of evidence (bi-variate analyses, retrospective analyses, etc.) Two recent prospective studies (Allam, 1998; Clark, 1999, personal communication) found that the key dynamic components of the SRA improved upon assessments using solely static factors.

Although Static-99 can meaningfully differentiate between sex offenders with higher or lower probabilities of recidivism, the labels used to describe the various risk levels (low, medium-low, medium-high, high) do not reflect any absolute standard of risk. The standard of tolerable risk depends on the context of the assessment. An offender with a 10% chance of sexual recidivism over 15 years may be an good candidate for conditional release (i.e., "low" risk), but an unacceptably high risk for holding positions of trust over children.

Conclusion

The present study is part of growing body of research supporting empirically based risk prediction for sexual offenders. No risk prediction scheme will be entirely accurate, and the measures described in the current article are far from perfect. Nevertheless, the current results are a serious challenge to sceptics who claim that sexual recidivism cannot be predicted with sufficient accuracy to be worthy of consideration in applied contexts. The value of unstructured clinical opinion can be questioned, but there is sufficient evidence to indicate that empirically based risk assessments can meaningfully predict the risk for sexual offence recidivism. It is up to future researchers and clinicians to build upon the foundations that have been already established.

References

Allam, J. (1998). Community-based treatment for sex offenders: An evaluation. Birmingham: University of Birmingham and West Midlands Probation Service.

Allison, P. D. (1984). Event history analysis: Regression for longitudinal event data. Beverly Hills, CA: Sage.

Becker, G. (1996). The meta-analysis of factor analysis: An illustration based on the cumulation of correlation matrices. Psychological Methods, 1, 341-353.

Bonta, J., & Hanson, R. K. (1994). Gauging the risk for violence: Measurement, impact and strategies for change. Ottawa: Ministry Secretariat, Solicitor General Canada.

Cortoni, F. A. (1998). The relationship between attachment styles, coping, the use of sex as a coping strategy, and juvenile sexual history in sexual offenders. Unpublished doctoral dissertation. Queen's University, Kingston, Ontario, Canada.

Doren, D. (1999, June). The accuracy of sex offender recidivism risk assessments. Presentation at the XXIV International Congress on Law and Mental Health, Toronto.

Epperson, D. L., Kaul, J. D., & Hesselton, D. (1998, October). Final report of the development of the Minnesota Sex Offender Screening Tool – Revised (MnSOST-R). Presentation at the 17th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Vancouver, B.C., Canada.

Freund, K., & Watson, R. (1991). Assessment of the sensitivity and specificity of a phallometric test: An update of phallometric diagnosis of pedophilia. Psychological Assessment, 3, 254-260.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adults offender recidivism: What works! Criminology, 34, 575-607.

Grubin, D. (1998). Sex offending against children: Understanding the risk. Police Research Series Paper 99. London: Home Office.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology, 143, 29-36.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under Receiver Operating Characteristic curves derived from the same cases. Radiology, 148, 839-843.

Hanson, R. K. (1997). The development of a brief actuarial risk scale for sexual offense recidivism. (User Report 97-04). Ottawa: Department of the Solicitor General of Canada.

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual

offender recidivism studies. Journal of Consulting and Clinical Psychology, 66 (2), 348-362.

Hanson, R. K., & Harris, A. J. R. (1998). Dynamic predictors of sexual recidivism. (User Report 1998-01). Ottawa: Department of the Solicitor General of Canada.

Hanson, R. K., & Harris, A. J. R. (in press). Where should we intervene? Dynamic predictors of sex offense recidivism. Criminal Justice and Behavior.

Hanson, R. K., Scott, H., & Steffy, R. A. (1995). A comparison of child molesters and non-sexual criminals: Risk predictors and long-term recidivism. Journal of Research in Crime and Delinquency, 32(3), 325-337.

Hanson, R. K., Steffy, R. A., & Gauthier, R. (1992). Long-term follow-up of child molesters: Risk prediction and treatment outcome. (User Report No. 1992-02.) Ottawa: Corrections Branch, Ministry of the Solicitor General of Canada.

Hanson, R. K., Steffy, R. A., & Gauthier, R. (1993a). Long-term recidivism of child molesters. Journal of Consulting and Clinical Psychology, 61, 646-652.

Hanson, R. K., Steffy, R. A., & Gauthier, R. (1993b). [Long-term recidivism of child molesters]. Unpublished raw data.

Hare, R. D. (1991). The Hare Psychopathy Checklist – Revised. Toronto, Ontario: Multi-Health Systems.

Hedges, L. V. (1994). Fixed effect models . In H. Cooper & L. V. Hedges (Eds.), The handbook of research synthesis (pp. 285-299). New York: Russell Sage Foundation.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. New York: Academic Press.

Hemphill, J. F., Hare, R. D., & Wong, S. (1998). Psychopathy and recidivism: A review. Legal and Criminological Psychology, 3, 139-170.

Janus, E. S., & Meehl, P. E. (1997). Assessing the legal standard for predictions of dangerousness in sex offender commitment proceedings. Psychology, Public Policy, and Law, 3, 33-64.

McClish, D. K. (1992). Combining and comparing area estimates across studies or strata. Medical Decision Making, 12, 274-279.

Metz, C. E. (1998). ROCKIT (Version 0.9.1). [Computer software]. Chicago, IL: University of Chicago.

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

Ozer, D. J. (1985). Correlation and the coefficient of determination. Psychological Bulletin, 97, 307-315.

Pellerin, B., Proulx, J., Ouimet, M., Paradis, Y., McKibben, A., & Aubut, J. (1996). Ètude de la récidive post-traitement chez des agresseirs sexuels judiciarisés. Criminologie, 29, 85-108.

Phenix, A., & Hanson, R. K. (in press). Coding rules for scoring the RRASOR. Thousand Oaks, CA: Sage.

Proulx, J., Pellerin, B., McKibben, A., Aubut, J., & Ouimet, M. (1997). Static and dynamic predictors of recidivism in sexual offenders. Sexual Abuse, 9, 7-28.

Proulx, J., Pellerin, B., McKibben, A., Aubut, J., & Ouimet, M. (1995). [Static and dynamic predictors of recidivism in sexual aggressors]. Unpublished raw data.

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., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. Journal of Interpersonal Violence, 10(1), 85-105.

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. (1996). [Recidivism information on 288 sexual offenders released from the Oakridge Mental Health Centre, Penetanguishene, Ontario]. Unpublished data set.

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 follow-up of rapists assessed in a maximum-security psychiatric facility. Journal of Interpersonal Violence, 5(4), 435-448.

Rice, M. E., Quinsey, V. L., & Harris, G. T. (1991). Sexual recidivism among child molesters released from a maximum security institution. Journal of Consulting and Clinical Psychology, 59, 381-386.

Seidman, B. T., Marshall, W. L., Hudson, S. M., & Robertson, P. J. (1994). An examination of intimacy and loneliness in sex offenders. Journal of Interpersonal Violence, 9, 518-534.

Soothill, K. L., & Gibbens, T. C. N. (1978). Recidivism of sexual offenders. British Journal of Criminology, 18, 267-276.

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. Psychological Bulletin, 99, 100-117.

Thornton, D. (1997). [A 16-year follow-up of 563 sexual offenders released from HM Prison Service in 1979.] Unpublished raw data.

Wilson, R. J. (1999). Emotional congruence in sexual offenders against children. Sexual Abuse: A Journal of Research and Treatment, 11, 33-47.

APPENDIX I

Coding rules of Static-99.

Risk Factor	Codes		Score
Prior Sex Offences (Same rules as in RRASOR)	Charges	Convictions	
	None	None	0
	1-2	1	1
	3-5	2-3	2
	6 +	4 +	3
Prior sentencing dates (excluding index)	3 or less		0
	4 or more		1
Any convictions for non-contact sex offences	No		0
	Yes		1
Index non-sexual violence	No		0
	Yes		1
Prior non-sexual violence	No		0
	Yes		1
Any Unrelated Victims	No		0
	Yes		1
Any Stranger Victims	No		0
	Yes		1
Any Male Victims	No		0
	Yes		1
Young	Aged 25 or older		0
	Aged 18 – 24.99		1
Single	Ever lived with lover for at least two years?		
	Yes		0
	No		1
Total Score	Add up scores from individual risk factors		

Notes

Static 99 is intended for males aged at least 18 who are known to have committed at least one sex offence.

1) Prior sex offences. Count only officially recorded offences. These could include a) arrests and charges, b) convictions, c) institutional rules violations, and d) probation, parole or conditional release violations arising from sexual assault, sexual abuse, sexual misconduct or violence engaged in for sexual gratification.

Non-sexual offences resulting from sexual behaviour would also be included as sexual offences (e.g., voyeur convicted of trespass by night). When the offence behaviour was sexual, but resulted in a conviction for a violent offence (e.g., assault, murder), then the offender is considered to have committed both a sexual and non-sexual violent offence and could receive points for both items.

Count only the number of sexual convictions or charges prior to the index offence. Do not count the sex offences included in the most recent court appearance. Institutional rule violations and conditional release violations count as one charge. Use either charges or convictions, whichever indicates the higher risk. More detailed worked examples of scoring prior offences are given in the RRASOR scoring guidelines (Phenix & Hanson, in press).

2) Prior sentencing dates. Count the number of distinct occasions on which the offender has been sentenced for criminal offences of any kind. The number of charges/convictions does not matter, only the number of sentencing dates. Court appearances that resulted in complete acquittal are not counted. The index sentencing date is not included.

3) Non-Contact Offences. This category includes convictions for non-contact sexual offences, such as exhibitionism, possessing obscene material, obscene telephone calls, and voyeurism. Self-reported offences do not count in this category.

4) Index Non-sexual Violence. Refers to convictions for non-sexual assault that are dealt with on the same sentencing occasion as the index sex offence. These convictions can involve the same victim as the index sex offence or they can involve a different victim. All non-sexual violence convictions are included providing they were dealt with on the same sentencing occasion as the index sex offences. Example offences would include murder, wounding, assault causing bodily harm, assault, robbery, pointing a firearm, arson, and threatening.

5) Prior Non-sexual Violence. The category includes any conviction for non-sexual violence prior to the index sentencing occasion.

The previous items (Items 1-5; prior offences) are based on officially records. The following items are based on all available information, including self-report, victim accounts, and collateral contacts.

6) Unrelated Victim. A related victim is one where the relationship would be sufficiently close that marriage would normally be prohibited, such as parent, uncle, grand-parent, step-sister.

7) Stranger Victim. A victim is considered to be a stranger if the victim did not know the offender 24 hours before the offence.

8) Male Victim. Included in this category are all sexual offences involving male victims. Possession of child pornography involving boys, however, would not count in this category.

9) Young. This item refers to the offender's age at the time of the risk assessment. If the

assessment concerns the offender's current risk level, it would be his current age. If the assessment concerns an anticipated exposure to risk (e.g., release, reduced security at some future date), the relevant age would be his age when exposed to risk.   Static-99 is not intended for those who are less than 18 years old at the time of exposure to risk.

10) Single. The offender is considered single if he has never lived with a lover (male or female) for at least two years. Legal marriages involving less than two years of co-habitation do not count.

TRANSLATING STATIC 99 SCORES INTO RISK CATEGORIES

Score	Label for Risk Category
0,1	Low
2,3	Medium-Low
4,5	Medium-High
6 plus	High