Law Hum Behav (2010) 34:198–211
DOI 10.1007/s10979-009-9180-1

ORIGINAL ARTICLE

Predicting Recidivism Amongst Sexual Offenders: A Multi-site Study of Static-2002

R. Karl Hanson · Leslie Helmus · David Thornton

Published online: 18 June 2009

Abstract The predictive accuracy of Static-2002 (Hanson & Thornton, Notes on the development of Static-2002 (Corrections Research User Report No. 2003-01), 2003) was examined in eight samples of sexual offenders (five Canadian, one U.S., one U.K., one Danish; total sample of 3,034). Static-2002 showed moderate ability to rank order the risk for sexual, violent and general (any) recidivism (AUCs of .68, .71, and .70, respectively), and was more accurate than Static-99. These findings support the use of Static-2002 in applied assessments. There were substantial differences across samples, however, in the observed sexual recidivism rates. These differences present new challenges to evaluators wishing to use actuarial risk scores to estimate absolute recidivism rates.

Keywords Sex offenders · Recidivism · Prediction · Static-2002

Actuarial risk tools are now routinely used in applied risk assessment with offenders (Archer, Buffington-Vollum, Stredny, & Handel, 2006). Such tools specify the factors to consider in the risk assessment, the method for combining the items into an overall score and the expected recidivism rates associated with the scores (Dawes, Faust, & Meehl, 1989). For the prediction of sexual recidivism, Static-99 (Hanson & Thornton, 2000) is by far the most commonly used actuarial risk tool in Canada and the U.S. for treatment planning (McGrath, Cumming, & Burchard, 2003), community supervision (Interstate Commission for Adult Offender Supervision, 2007) and civil commitment evaluations (Jackson & Hess, 2007). As well, it is used in jurisdictions as diverse as Sweden, Belgium, Israel, Singapore and Japan. Static-99 is also the most researched of all risk assessment tools for sex offenders, with moderate predictive accuracy (on average) amongst 63 replication studies (Hanson & Morton-Bourgon, in press).

Static-99 contains 10 items covering static, historical factors (such as age and prior offences) and can be reliably scored without advanced professional training. More complex risk assessment systems are available that may be more accurate than Static-99 for predicting sexual recidivism (e.g., Olver, Wong, Nicholaichuk, & Gordon, 2007; Thornton, 2002), and violent recidivism (e.g., G. T. Harris et al., 2003); however, the popularity of Static-99 demonstrates the widespread demand for cost-effective risk tools applicable to a wide range of sexual offenders.

Hanson and Thornton (2003) created Static-2002 as a potential improvement over Static-99. It was designed to have the same basic features of Static-99, namely, a brief actuarial measure for the prediction of sexual recidivism based on commonly available information; however, it was hoped that Static-2002 could address some of the weaknesses of Static-99.

Static-99 was created by merging two previously existing scales (RRASOR and Structured Anchored Clinical Judgement—SACJ-Min), which resulted in different definitions for different items (e.g., charges count for sexual offences, whereas only convictions count for non-sexual violence). With Static-2002, the authors attempted to standardize the coding rules by selecting the definitions with the strongest support in pilot studies. Unlike Static-99, the Static-2002

R. K. Hanson (✉) · L. Helmus
Corrections Research, Public Safety Canada, 340 Laurier Ave, West, Ottawa, Ontario K1A 0P8, Canada
e-mail: karl.hanson@ps-sp.gc.ca

D. Thornton
Sand Ridge Secure Treatment Centre, Mauston, WI, USA

items were organized into subscales to aid interpretation. It was also hoped that Static-2002 would be more accurate than Static-99 at predicting sexual and violent recidivism.

Most of the previous research on the accuracy of Static-99 and Static-2002 has focussed on the ability of these risk tools to differentiate offenders on their risk for recidivism. For example, predictive accuracy is routinely reported in terms of correlation coefficients, areas under receiver operating characteristics curves (AUC for ROC), or standardized mean differences (Cohen's *d*). These indices describe the extent to which the recidivists are different from the non-recidivists, but provide no information about the absolute recidivism rates. Even when the AUCs are consistent across studies, it is possible for there to be meaningful differences in the observed recidivism rates (Mossman, 2006). Relatively little research has examined the stability of the observed recidivism rates for actuarial risk tools for sexual offenders (see Doren, 2004a, for an exception), and conventions have yet to be developed concerning the best ways to report predictive accuracy in terms of absolute recidivism rates.

At a broad level, observed recidivism rates are a function of the factors included in the scale as well as factors not measured in the scale (plus error). Some of the factors not included in the risk tool should be (at least for a comprehensive evaluation of risk): namely, individual risk factors that add incrementally to those already included in the scale. Other factors external to the scale involve relatively arbitrary decisions concerning research design and recidivism definitions (e.g., year of release, length of follow-up). The offenders' environment would also influence recidivism rates (e.g., treatment, supervised release). As well, there is likely to be variation in the extent to which new offences are detectable (e.g., willingness of victims to report, police efficiency). When samples systematically differ on these external (unmeasured) factors, different *base rates* of recidivism would be expected. These differences would persist even when the outcome is defined consistently for identical risk scores. Consequently, it is useful to separate base rates from discriminative accuracy when examining the stability of absolute recidivism rates.

The purpose of the present study is to compare the ability of Static-99 and Static-2002 to rank sexual offenders according to relative recidivism risk in diverse samples, and to assess the ability of Static-2002 to predict observed recidivism rates at 5 and 10 years post release. Logistic regression was used as the primary method of estimating absolute recidivism rates (Hosmer & Lemeshow, 2000, § 3.9). One advantage of logistic regression is that it fits a constant term (B_0), which estimates the recidivism base rate and B_1, which estimates the average change in recidivism rates for adjacent scores of the risk tool (discriminative accuracy).

In the original development study (Hanson & Thornton, 2003), Static-2002 was slightly better than Static-99 for predicting sexual recidivism (AUC of .71 and .70, respectively) and violent recidivism (AUC of .71 and .69). There was considerable missing data in the developmental samples, however, and the authors recommended that further research be conducted before Static-2002 is used in applied assessments. That research has now been completed. The current study summarizes information on the predictive accuracy of Static-2002 by analyzing a dataset created from all known Static-2002 recidivism studies.

METHOD

Measures

Static-2002 (Hanson & Thornton, 2003)

Static-2002 is a 14-item[1] actuarial measure that assesses recidivism risk of adult male sexual offenders. The items are organized into five subscales: age (at release), persistence of sexual offending (prior sentencing occasions for sexual offences, any juvenile arrest for a sexual offence, rate of sexual offending), deviant sexual interests (any non-contact sex offences, any male victim, young/unrelated victims), relationship to victims (any unrelated victim, any stranger victim), and general criminality (any prior involvement with the criminal justice system, prior sentencing occasions, any community supervision violation, years free prior to index sex offence, any prior non-sexual violence). Offenders can be placed in one of five risk categories based on their total score (ranging from 0 to 14): low (0–2), low–moderate (3, 4), moderate (5, 6), moderate–high (7, 8) and high (9 + ; see Helmus, 2007).

Static-99 (Hanson & Thornton, 2000)

Static-99 is a 10-item actuarial measure that assesses recidivism risk of adult male sexual offenders. Offenders can be placed in one of four risk categories based on their total score (ranging from 0 to 12): low (0, 1), moderate–low (2, 3), moderate–high (4, 5) and high (6 +). For further information, see http://www.static99.org.

[1] Static-2002 originally had 13 items (Hanson & Thornton, 2003). To increase clarity in the coding manual (Phenix, Doren, Helmus, Hanson, & Thornton, 2008), one item was divided into two separate items. This does not change the total score (the summed score of the two separate items in the new coding rules is identical to the score for the old item).

Samples

To be included, samples required Static-99 scores, Static-2002 scores and information on sexual recidivism. Raw datasets from nine samples were obtained, representing all known Static-2002 replications conducted as of December 2006. Prior to merging, each dataset was cleaned by checking for internal inconsistencies (e.g., miscalculation of total scores or divergent scoring of identical items across Static-99 and Static-2002). Identified errors were corrected if possible; otherwise, the case was deleted. Cases were also deleted if more than one Static-2002 item was missing or any Static-99 item other than co-habitation (Item 2) was missing. Consequently, the sample sizes reported in the current study were not identical to those reported in other publications using these samples.

Datasets varied greatly in the number of errors detected. One dataset was not used due to a high number of errors. In the seven remaining datasets (one did not include individual item scores and therefore could not be checked), the proportion of cases with a detected coding error ranged from 1.6% to 18.9%, with a median of 7.1%. For the sample with an error rate of 18.9%, 86% of those errors were attributed to incorrectly computing the Static-2002 total score. Other errors included coding offenders as having prior sentencing occasions for sex offences but not for general offending (which includes sexual offences) or coding an offender as having a stranger victim but not an unrelated victim (not possible according to the coding rules).

Table 1 displays characteristics of the eight samples retained for analysis ($n = 3{,}034$, all male). Five samples were Canadian and there was one each from the U.S., U.K. and Denmark. Three of the studies examined relatively representative (unselected) samples of offenders from the Correctional Service of Canada (CSC), which administers federal sentences (two years or more; Bigras, 2007; Boer, 2003; Langton et al., 2007). Three samples were pre-selected to be high risk (Bengtson, 2008; Haag, 2005; Knight & Thornton, 2007), one was a relatively low-risk community sample (Hanson, Harris, Scott, & Helmus, 2007), and one study examined offenders from a variety of settings (Harkins & Beech, 2007).

Most offenders were released from institutional settings ($k = 6$). Based on the demographics of the correctional populations from the countries sampled, it can be assumed that most offenders were Caucasian. All but one of the samples included information on sexual, violent (including sexual) and any recidivism (the remaining sample included sexual recidivism only). Of the seven samples that provided individual item scores ($n = 2{,}549$), 47 cases (1.8%) had missing information on one item for Static-99 or Static-2002. Missing items were coded zero as recommended by the coding manuals (A. J. R. Harris et al., 2003; Phenix et al., 2008). Each sample will be briefly described (further information is available upon request).

CSC: B.C. (Boer, 2003)

This sample included all male federal sex offenders in B.C. whose Warrant Expiry Date (WED; the end of their sentence) was between January 1990 and May 1994 and who had sufficient information to code the measures. Exposure to treatment was unknown but during the time period when these offenders were incarcerated, CSC offered sexual offender treatment programmes to all sex offenders. Offenders were classified as child molesters if the majority of their victims were 12 years old or younger, and rapists if

Table 1 Sample characteristics

Sample	*n*	Age *(SD)*	Offender type: Rape/CM (%)	Average follow-up (Years)	Recidivism rate				Static-99 *M (SD)*	Static-2002 *M (SD)*
					Sexual	Any violent	Any	Outcome criteria		
CSC: B.C.	299	41 (13)	41/55	13.3	8.7	23.4	48.5	Conviction	3.3 (2.3)	4.6 (2.5)
CSC: Quebec	487	40 (12)	37/46	4.6	6.4	15.4	24.0	Charge	2.7 (2.0)	4.3 (2.4)
CSC: Warkworth	354	40 (11)	37/53	5.8	11.0	24.3	39.5	Conviction	3.6 (2.1)	5.4 (2.3)
CSC: Detained	198	37 (10)	60/36	7.0	25.3	–	–	Conviction	3.9 (2.0)	5.9 (2.2)
Bridgewater: MTC (MA)	485	36 (12)	42/58	8.6	25.8	36.5	52.6	Charge	4.5 (2.2)	6.6 (2.5)
Denmark psychiatric	311	33 (10)	50/49	14.8	33.8	51.8	64.6	Charge	3.8 (2.1)	5.1 (2.3)
DSP (Canada)	702	42 (13)	36/50	3.4	8.1	16.4	27.9	Charge	2.9 (2.0)	4.1 (2.3)
U.K. Multi-site treatment	198	43 (13)	10/82	10.4	14.1	20.7	35.9	Conviction	2.8 (2.2)	4.3 (2.6)
Total	3,034	39 (12)	39/53	7.5	15.2	25.6	39.7		3.4 (2.2)	5.0 (2.6)

Note: CSC = Correctional Service Canada. Age = Age at release

the majority of their victims were 17 years old and above. For offenders with victims between ages 13–16, a judgment was made based on the offence circumstances or the offender was classified as mixed.

Recidivism information was collected using Canadian Police Information Centre (CPIC) records, maintained by the Royal Canadian Mounted Police (RCMP). CPIC records contain basic criminal history information: date of conviction, offence title (according to the Canadian Criminal Code), and sentence. Information on charges that were stayed or for which the offender received an acquittal are inconsistently recorded on CPIC records, and offence details are not recorded. Category B sexual offences (see A. J. R. Harris et al., 2003) were excluded from the definition of sexual recidivism. As Category B offences are uncommon, their exclusion would likely have a trivial impact on the results.

CSC: Quebec (Bigras, 2007)

This sample contained sexual offenders given a federal sentence in Quebec between 1995 and 2000. Exposure to treatment was unknown but CSC would have offered treatment to all offenders. Offenders were classified as child molesters if their victims were 12 years old or younger. Offenders with victims age 16 and above were classified as rapists. Offenders with victims between 13 and 15 years old were classified as child molesters if the victim was related, and rapists if the victim was unrelated. Offenders with victims in multiple categories were classified as mixed. Recidivism data was collected using CPIC records.

CSC: Warkworth (Langton et al., 2007)

This study followed sex offenders offered treatment at Warkworth Sexual Behaviour Clinic (WSBC; a medium security institution) between 1989 and 2001. Most (86%) of the offenders completed treatment, 8% dropped out and 6% refused treatment. Offender type was unavailable, although the overall proportion of rapists and child molesters in this sample is reported elsewhere (Langton et al., 2007). Recidivism information was coded from CPIC records.

CSC: Detained (Haag, 2005)

This sample included all male federal sex offenders whose WED was in 1995. In the original sample, most offenders (75%) were in the community on some form of release prior to their WED, but recidivism information only started at their WED. To obtain complete follow-up data, only offenders detained until Warrant Expiry were retained in the current study. Under Canadian legislation, offenders are to be automatically released after serving two-thirds of their sentence. Offenders may be detained if the parole board is satisfied that the offender poses a significant risk of committing a serious offence before their sentence expires.

Half of the sample (52%) received sex offender treatment in prison, 11% dropped out of treatment and 37% did not receive any treatment. The high refusal and dropout rates would be expected in a detained sample because these factors would likely influence the decision to detain. Recidivism information was collected from CPIC records and only sexual recidivism was coded. Offenders were classified as child molesters if all their victims were 13 years old or younger. Offenders with victims age 14 and above were classified as rapists. Offenders with victims in both categories were not classified.

Bridgewater: Massachusetts Treatment Center (MTC; Knight & Thornton, 2007)

This sample included offenders who were either assessed or treated at MTC between 1959 and 1984. MTC is a treatment centre for sexually dangerous persons. Although approximately half the sample was treated at the MTC, this treatment would not be consistent with contemporary standards. Recidivism information was obtained from four sources: Massachusetts Board of Probation records, Massachusetts Parole Board records, MTC Authorized Absence Programme records and FBI records. Offenders were classified as child molesters if all their victims were less than 16 years old, and rapists if all their victims were 16 or older. Offenders with victims in both categories were classified as mixed. Two hundred and thirty-two Static-99 cases and 258 Static-2002 cases were coded by two raters. When they disagreed, the average of the two discrepant scores was entered. As this was the only dataset where risk scores could include a fraction, scores were rounded. Rounding was done to the nearest even number to prevent artificial inflation of scores.

Denmark: Psychiatric (Bengtson, 2008)

This study followed sex offenders who received a pre-trial forensic psychiatric evaluation between 1978 and 1992 at one of two settings in Denmark. Such evaluations were typically conducted for offenders deemed high risk by the courts, suspected of mental disorder, accused of serious offences and those for whom an indefinite sentence was being considered. Exposure to treatment was unknown. Recidivism information was obtained from the Danish Central Crime Register. Offender type was unavailable in this dataset, although the overall proportion of rapists and

child molesters in this sample is reported elsewhere (Bengtson, 2008).

Canada: Dynamic Supervision Project (Hanson et al., 2007)

This prospective study followed offenders on community supervision between 2001 and 2005 in Canada, Alaska and Iowa. However, only Canadian offenders had sufficient information to score Static-2002. Approximately half of the offenders served custodial sentences prior to being on community supervision. Exposure to treatment was unknown. Recidivism information was obtained from several sources: probation officers, CPIC records and police jurisdictions. Offenders were classified as child molesters if all their victims were 13 years old or younger, or if they had only related victims less than 18 years old. Offenders were classified as rapists if they had unrelated victims age 14 and above or any victim age 18 or older. Offenders with victims in both categories were classified as mixed.

U.K.: Multi-site Treatment (Harkins & Beech, 2007)

This sample included offenders from three other studies in the United Kingdom. The first sample consisted of offenders from the mid to late 1990s in Her Majesty's Prison Service's Sex Offender Treatment Programme. The second sample was drawn from a community treatment programme in the early 1990s, and the third sample included offenders who underwent community treatment in the late 1990s in West Midlands. The intensity of treatment varied and dropouts were retained in the sample. Recidivism data was collected from the Home Office Offenders Index and Police National Computer. Offender type was not available in this dataset, although the overall proportion of rapists and child molesters in this sample is reported elsewhere (Harkins & Beech, 2007).

Overview of Analyses

The first set of analyses used areas under receiver operating characteristic curves (ROC AUC) to compare the ability of Static-99 and Static-2002 to rank order offenders in terms of risk to reoffend. The AUC is useful for comparing results across samples because it is not influenced by recidivism base rates (Rice & Harris, 1995). It is, however, influenced by the variance in the scores used to predict recidivism (Hanson, 2008; Humphreys & Swets, 1991). To test whether Static-99 and Static-2002 differed in their level of predictive accuracy, Hanley and McNeil's (1983) test of correlated ROC areas was used.

Findings across studies were aggregated using fixed-effect meta-analysis (Hedges, 1994). To test the variability of effects across studies, Cochran's Q statistic was used (Hedges & Olkin, 1985). A significant Q statistic indicates that there is more variability across studies than would be expected by chance. Given significant variation across studies in the AUCs, the second set of analyses used logistic regression (Hosmer & Lemeshow, 2000) and Cox regression (Allison, 1984) to identify the source of this variation. Cox regression estimates relative risk ratios (hazard rates) associated with one or more predictor variables from survival data with unequal follow-up times.

The final set of analyses examined the 5 and 10 year recidivism rates associated with each Static-2002 score. Three methods were used to estimate absolute recidivism rates: (a) life table survival analysis (Soothill & Gibbens, 1978), (b) observed recidivism rates for fixed follow-up periods and (c) predicted values from the fitted (smoothed) logistic regression curves. All analyses were calculated using SPSS version 15, except the standard error of the logistic regression estimates, which were calculated using Equation 11.9 from Fleiss, Levin, and Paik (2003).

RESULTS

Across the eight samples, the AUC areas for the prediction of sexual, violent and any recidivism were all statistically significant (the 95% confidence intervals did not include .5), ranging from .59 to .77 for Static-99 and from .64 to .79 for Static-2002 (see Table 2). There was more variability across samples than would be expected by chance (all Q statistics were statistically significant for both Static-99 and Static-2002—See Table 3). The meta-analysis of the individual samples (Table 3) found that Static-2002 had consistently higher AUC areas than Static-99 for sexual recidivism (.68 vs. .66), violent (including sexual) recidivism (.70 vs. .66) and any recidivism (.71 vs. .66). The 95% confidence intervals for the difference between measures did not include zero, meaning that the difference between Static-99 and Static-2002 was statistically significant for all outcomes. For these comparisons, all Q statistics were nonsignificant, indicating that the relative difference in predictive accuracy between Static-99 and Static-2002 did not vary significantly across samples. The remaining analyses looked at Static-2002 only.

Logistic regression was used to examine the extent to which the variation across samples could be attributed to different recidivism base rates. With one predictor variable, B_0 is a measure of the recidivism base rate for the sample (technically, the log odds of the predicted recidivism rate for offenders with a score of zero) and B_1 is an

Table 2 Predictive accuracy of Static-99 and Static-2002 in eight samples

Sample	*n*	Sexual recidivism		Violent recidivism		Any recidivism	
		AUC	95% C.I.	AUC	95% C.I.	AUC	95% C.I.
CSC: B.C.							
Static-99	299	.73	(.61, .84)	.69	(.62, .76)	.72	(.66, .77)
Static-2002	299	.71	(.61, .82)	.72	(.66, .79)	.78	(.73, .83)
CSC: Quebec							
Static-99	487	.63	(.52, .74)	.63	(.57, .69)	.63	(.57, .68)
Static-2002	487	.66	(.56, .76)	.67	(.61, .73)	.68	(.63, .74)
CSC: Warkworth							
Static-99	354	.61	(.52, .70)	.61	(.54, .67)	.63	(.57, .69)
Static-2002	354	.68	(.60, .76)	.68	(.62, .74)	.68	(.62, .73)
CSC: Detained							
Static-99	198	.66	(.57, .75)	–	–	–	–
Static-2002	198	.66	(.57, .74)	–	–	–	–
Bridgewater: MTC							
Static-99	485	.61	(.56, .67)	.62	(.56, .66)	.59	(.54, .64)
Static-2002	485	.64	(.58, .69)	.64	(.59, .69)	.64	(.59, .69)
Denmark Psychiatric							
Static-99	311	.61	(.55, .68)	.65	(.59, .71)	.63	(.56, .70)
Static-2002	311	.64	(.57, .70)	.67	(.61, .73)	.67	(.60, .73)
DSP							
Static-99	702	.76	(.70, .83)	.73	(.68, .78)	.72	(.68, .76)
Static-2002	702	.76	(.70, .82)	.76	(.72, .81)	.76	(.72, .80)
U.K. Multi-site treatment							
Static-99	198	.77	(.68, .86)	.73	(.65, .81)	.73	(.65, .80)
Static-2002	198	.79	(.70, .88)	.77	(.69, .85)	.76	(.69, .83)

Table 3 Meta-analysis of prediction AUCs for Static-99 and Static-2002

Recidivism outcome	Measure	Weighted AUC	95% Confidence interval		*k*	*n*	*Q*
Sexual							
	Static-99	.665	.638	.692	8	3,034	23.19*
	Static-2002	.685	.658	.711	8	3,034	16.46*
Violent							
	Static-99	.662	.639	.685	7	2,836	17.74*
	Static-2002	.702	.681	.724	7	2,836	19.92*
Any							
	Static-99	.664	.643	.684	7	2,836	24.62*
	Static-2002	.714	.694	.733	7	2,836	26.79*
Difference between Static-99 and Static-2002							
Sexual		.0183	.0044	.0323	8	3,034	8.55
Violent		.0358	.0242	.0474	7	2,836	6.16
Any		.0470	.0365	.0575	7	2,836	4.03

* $p < .05$, on a Chi-square distribution with $k - 1$ degrees of freedom

estimate of the average change in recidivism rates for adjacent Static-2002 scores (expressed as log odds ratios). Logistic regression equations were fitted for the individual samples based on recidivism after fixed 5 year and 10 year follow-up periods. The results for samples with at least 10 recidivists were summarized using fixed-effect

Table 4 Meta-analysis of logistic regression coefficients for Static-2002 for recidivism within fixed 5 year and 10 year follow-up periods

	k	*n* recid/*N*	B_0	95% Confidence interval	*Q*	B_1	95% Confidence interval	Q
Sexual recidivism								
5 year	7	282/1,892	−3.14	−3.53, −2.74	17.62**	.255	.198, .313	5.69
10 year	4	236/1,085	−2.59	−3.02, −2.15	12.62**	.237	.174, .301	4.45
Violent recidivism								
5 year	6	414/1,701	−2.67	−3.00, −2.34	11.08*	.270	.220, .321	4.45
10 year	5	392/1,142	−2.27	-2.63, −1.90	12.73*	.281	.224, .338	8.74
Any recidivism								
5 year	6	663/1,704	−2.21	−2.49, −1.92	11.39*	.316	.269, .363	11.75*
10 year	5	586/1,149	−1.67	−2.01, −1.35	17.78**	.310	.254, .365	15.33**

* $p < .05$, ** $p < .01$

Note: k = number of samples; N = total sample size. Degrees of freedom for Q are $k - 1$

meta-analysis. Restricting the analyses to cases with 5 and 10 year follow-up periods reduced the overall sample size considerably (e.g., from a total of 3,034 to 1,923 for fixed 5 years).

As can be seen in Table 4, there were significant differences in the estimates of the recidivism base rate (B_0) across 5 year and 10 year time periods for sexual, violent and any recidivism (all Q's were significant). In contrast, the variability in the predictive accuracy (B_1) was no more than would be expected by chance for sexual and violent recidivism (Q's were nonsignificant). For the prediction of any recidivism, there was significant variability between samples in both base rates and predictive accuracy.

Given that the ROC and the logistic regression analyses both indicated significant differences in the properties of Static-2002 across samples, the next set of analyses used Cox regression to examine potential sources of these differences. After controlling for Static-2002 scores, Cox regression found that the recidivism rates were significantly higher in the samples preselected to be high risk (Bridgewater, Denmark Psychiatric, and CSC Detained) than in the three routine CSC samples (B.C., Quebec, and Warkworth) for sexual recidivism (χ^2 change = 50.61, $df = 1$, $p < .001$; $k = 6$; $n = 2{,}134$) and violent recidivism (χ^2 change = 14.38, $df = 1$, $p < .01$; $k = 5$, $n = 1{,}936$—note that one of the preselected high risk samples [CSC Detained] examined only sexual recidivism). The two preselected high risk samples were not at a significantly greater risk for any recidivism than the three routine CSC samples (χ^2 change = 2.16, $df = 1$, $p > .05$; $k = 5$, $n = 1{,}936$). For all three outcomes, however, there were significant interactions between Static-2002 and sample (for sexual recidivism, χ^2 change = 6.98, $df = 1$, $p < .01$; for violent recidivism, χ^2 change = 12.56, $df = 1$, $p < .01$; for any recidivism, χ^2 change = 19.10, $df = 1$, $p < .01$). In all cases, the direction of the interaction effects was such that Static-2002 showed larger changes in the rate ratios for each additional point on Static-2002 in the routine CSC samples than in the samples preselected to be high risk. Figure 1 depicts the difference in estimated 10 year sexual recidivism rates per score on Static-2002 for the preselected high risk and routine CSC samples (the analyses to generate these estimates will be described in further detail below).

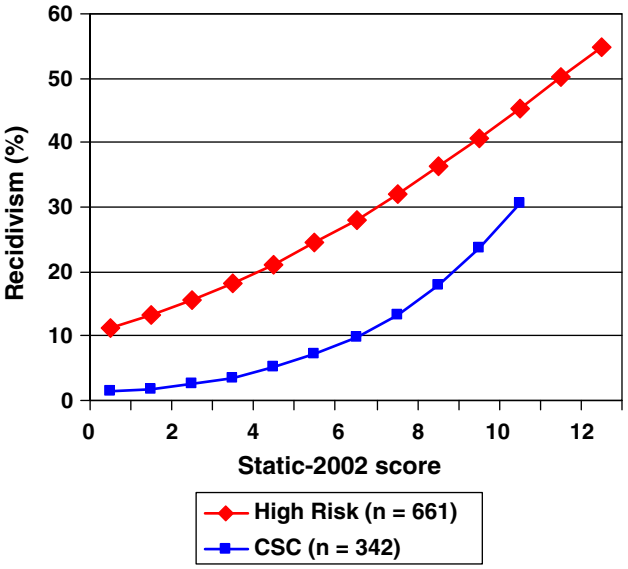

Fig. 1 Ten-year sexual recidivism rates generated by logistic regression

The next set of analyses examined the extent to which Static-2002 provided similar results for rapists and child molesters. Four datasets contained victim information with

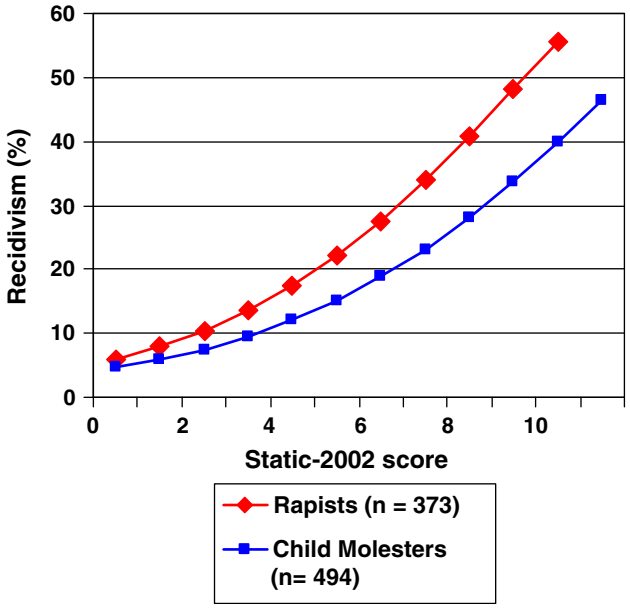

Fig. 2 Five-year violent recidivism rates generated by logistic regression

all three outcomes (DSP, Bridgewater, CSC Quebec and CSC B.C.) and one additional sample contained victim information and sexual recidivism only (CSC Detained). In these five datasets, there were 835 rapists and 1,025 child molesters. The classification followed the definitions used by the original researchers and mixed offenders were excluded from these analyses.

After controlling for Static-2002, Cox regression analyses found significantly higher recidivism rates amongst rapists for sexual recidivism (χ^2 change = 4.51, $df = 1$, $p < .05$; $k = 5$, $n = 1{,}860$), violent recidivism (χ^2 change = 31.87, $df = 1$, $p < .001$; $k = 4$; $n = 1{,}671$), and any recidivism (χ^2 change = 56.63, $df = 1$, $p < .001$; $k = 4$; $n = 1{,}671$). Figure 2 depicts the difference in estimated 5 year violent recidivism rates per score on Static-2002 for rapists and child molesters. None of the interactions between Static-2002 and victim type were statistically significant (all χ^2 change < 0.20), suggesting that Static-2002 was equally good at ranking relative risk for both rapists and child molesters. There were insufficient unique samples to meaningfully test the interactions between sample (high risk/routine CSC) and victim type (rapist/child molester), or the three-way interaction with Static-2002.

Two methodological features related to the coding of recidivism information were also tested. One variable was whether charges or convictions were used as the criteria for recidivism. After controlling for Static-2002 scores, samples using charges had significantly higher rates of sexual recidivism (χ^2 change = 9.90, $df = 1$, $p < .01$; $k = 8$, $n = 3{,}034$) and violent recidivism (χ^2 change = 16.86, $df = 1$, $p < .001$; $k = 7$; $n = 2{,}836$). For any recidivism, the difference approached significance (χ^2 change = 3.508, $df = 1$, $p = .061$; $k = 7$; $n = 2{,}836$).

The second variable concerned the sources used to measure recidivism. Two samples (Dynamic Supervision Project and Bridgewater) used multiple and diverse sources (e.g., more than official criminal records) to obtain recidivism information, whereas the remaining samples used basic criminal records. After controlling for Static-2002 scores, the use of more comprehensive sources was not significantly related to sexual, violent or any recidivism (all χ^2 change < 0.30).

The next set of Cox regression analyses examined the relative contribution of the three moderator variables that were significantly associated with recidivism in previous analyses: sample type (preselected high risk/routine CSC), offender type (rapist/child molester) and recidivism definition (charges/convictions).

For sexual recidivism, all three moderator variables were significant when controlling for Static-2002. Sample type had the largest effect, with preselected high risk samples showing approximately 2.3 times the sexual recidivism rate of routine CSC samples (for comparison, offender type and recidivism definition had rate ratios of 1.3 and 1.4, respectively). When all three moderator variables were entered together in a model that controlled for Static-2002 scores, type of sample was the only variable that remained significant, and its rate ratio increased slightly to 2.5 (if entered last, χ^2 change = 26.61, $df = 1$, $p < .001$; $k = 4$; $n = 1{,}279$). The rate ratio for offender type decreased slightly (from 1.3 to 1.2) and was no longer significant (if entered last, χ^2 change = 1.61, $df = 1$, $p = .20$). The rate ratio for recidivism definition was no longer significant and changed direction, with samples using charges showing approximately 79% the recidivism rate of samples using convictions (if entered last, χ^2 change = 2.23, $df = 1$, $p = .14$).

For violent recidivism, all three moderator variables were initially significant in predicting recidivism above and beyond Static-2002. Offender type had the largest effect, with rapists showing 1.8 times more violent recidivism than child molesters (the rate ratios for sample type and recidivism definition were both 1.4). When all three moderator variables were entered in a model that controlled for Static-2002 scores, only offender type remained significant, with a rate ratio of 1.7 (if entered last, χ^2 change = 19.66, $df = 1$, $p < .001$; $k = 3$; $n = 1{,}090$). The rate ratios for sample type and recidivism definition were both nonsignificant (for sample type, rate ratio = 1.1, χ^2 change = .30, $df = 1$, $p = .58$; for recidivism definition, rate ratio = 1.2, χ^2 change = .87, $df = 1$, $p = .35$).

For any recidivism, offender type was the only moderator variable that was significant after controlling for Static-2002 scores, with rapists showing 1.8 times the general recidivism rate as child molesters. The other moderator variables were therefore not considered further.

Estimated Recidivism Rates

The above analyses indicate the potential value of separate prediction tables for rapists, child molesters, offenders preselected to be high risk and routine CSC samples. The recidivism rates are calculated using three methods: (a) life table survival analysis (Soothill & Gibbens, 1978), which is how the recidivism rates for Static-99 (A. J. R. Harris et al., 2003 were calculated; (b) observed recidivism rates for fixed follow-up periods and (c) predicted values from the fitted (smoothed) logistic regression curves (Hosmer & Lemeshow, 2000). Of the three estimates, we believe that the most confidence should be placed in the logistic regression estimates—provided that the logistic distribution is an acceptable fit to the data. An important advantage of the logistic regression estimates is that they represent a smoothed distribution based on the full dataset that has fixed follow-up periods. In comparison, the other procedures only use offenders with a specific score, which can lead to large random fluctuations when there are only a few offenders with a specific score. As an example, Fig. 3 depicts the 10 year sexual recidivism estimates derived from the complete sample using each of these three methods, with logistic regression correcting for the fluctuations evident in the other two methods.

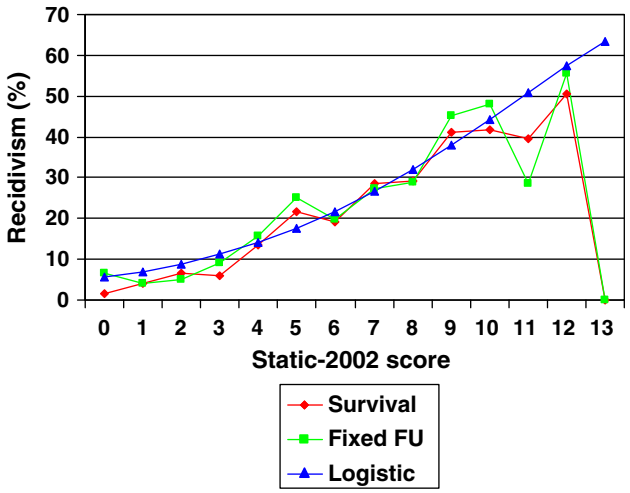

Fig. 3 Ten-year sexual recidivism rates generated from survival analysis, fixed follow-up periods and logistic regression

Logistic regression estimates are only accurate, however, when the observed rates approximate a logistic distribution. Based on the Hosmer–Lemeshow goodness-of-fit test, the logistic distribution was an acceptable approximation for all the sexual recidivism rates in this study (all chi-squares were nonsignificant). The distributions of violent and overall (any) recidivism rates, however, were significantly different from logistic in certain analyses (see Appendix, Table 5). When the logistic distribution is not an adequate fit, then preference should be given to the estimates from survival analysis, as they were based on more offenders than were the rates for the fixed follow-up periods.

A selection of recidivism rate tables is provided in Appendix, Tables 6 and 7; a larger selection of recidivism rate tables is available upon request.

DISCUSSION

The current study found that Static-2002 predicted all types of recidivism with significantly greater accuracy than Static-99. The differences in predictive accuracy, however, were not large—particularly for sexual recidivism (average AUC of .68 for Static-2002 vs. .66 for Static-99). The study also found that the predictive accuracy of both Static-99 and Static-2002 varied significantly across studies. When logistic regression analysis was used, the predictive accuracy of Static-2002 no longer showed significant variability but there was more variation than would be expected by chance in the observed recidivism rates per Static-2002 risk score. Post hoc analyses found that the recidivism rates were higher in samples preselected to be high risk compared to routine samples from CSC. As well, rapists showed higher recidivism rates than child molesters for equivalent risk scores.

The large sample sizes in the current study (mostly 3,000+) provided sufficient statistical power to detect even small differences. Consequently, it is important to consider the magnitude of the observed differences. Examination of the sexual recidivism rates for the preselected high risk samples and the routine CSC samples suggests that the differences would be large enough to be of interest to decision-makers (see Appendix, Table 7). Given a Static-2002 score of 4, for example, the 10 year sexual recidivism rate was 5.1% for the CSC samples compared to 21.2% for the high risk samples.

In contrast, the differences in the sexual recidivism rates between rapists and child molesters were minimal and would have little practical consequences in most contexts. For scores 0 through 8, the differences were less than 3% in

absolute value (e.g., 12.6% vs. 9.5%). For offenders with a score of 9 or more, the pattern actually reversed, such that the sexual recidivism rates of child molesters were marginally higher than for rapists. For violent recidivism, the rates for rapists tended to be approximately 10% higher than the rates for child molesters, which may be of interest for certain decisions. These tables were not included but are available upon request.

The observation that the recidivism rates for specific Static-2002 scores varies across samples raises important questions about how Static-2002 and other actuarial risk tools should be used in applied assessment. Until now, all the actuarial risk tools for sexual offenders have had only one prediction table, implying that the expected recidivism rates are stable. Any deviations from the group averages were based on secondary considerations and weaker evidence (typically some form of post hoc adjustment or untested professional judgement).

The current results demonstrate that variation in recidivism rates across samples cannot be ignored—even by the most empirically minded evaluators. Given a score of 5, for example, the 95% confidence intervals for the predicted 10-year sexual recidivism rate for the total sample of offenders (15.4% to 20.3%) did not overlap with the confidence intervals for the routine CSC offenders (4.6% to 10.7%), nor with the confidence intervals for the preselected high risk offenders (21.0% to 28.4%). Although these findings only concern Static-2002, there is no reason to believe that other actuarial tools are exempt. To our knowledge, Static-2002 is the only risk tool for sexual offenders for which the recidivism rates across large, diverse samples have been examined using individual case data, consistent follow-up times and statistics that separate base rate estimates from estimates of relative risk. Ongoing research on Static-99 suggests broadly similar variations across samples (Harris, Helmus, Hanson, & Thornton, 2008).

Evaluators may be tempted to hide the differences in recidivism rates across samples by only interpreting the recidivism rates for the total sample. The recidivism rates for the total sample, however, are only meaningful if the population from which the sample is drawn can be clearly specified. This is not the case. The characteristics of the total sample depended on the data available, and, in our study, included a larger proportion of high risk offenders than would be found in representative samples of sexual offenders. Rather than interpreting an arbitrary collection of samples, it would be more defensible for evaluators to recognize meaningful variation across samples and report the range of recidivism rates observed.

The observed range, however, may be insufficiently precise for certain decisions. In particular, many civil commitment statutes require a finding that the offender is more likely than not to reoffend sexually. This threshold can be interpreted as observed recidivism rates of 51%, although lower figures are frequently used in order to make allowance for extended follow-up and undetected offending. As can be seen in Fig. 1, only offenders preselected to be high risk would be expected to come near an expected recidivism rate of 50%. The highest sexual recidivism rate for the routine CSC cases was only 30.5% after 10 years. Consequently, evaluators wishing to use Static-2002 in civil commitment evaluation would need to consider more than the score in order to make a finding of "more likely than not" to reoffend sexually. As research progresses, the variables responsible for the differences between samples will be identified and included in new actuarial risk tools. Until that time, evaluators will be forced to use their professional judgement if they want to restrict an offender's risk to a range narrower than that bounded by the routine CSC and preselected high risk samples.

Understanding potential differences between the preselected high risk and routine CSC samples will be helpful to both researchers and evaluators. The preselected high risk samples in the current study had been determined by a court or an administrative tribunal to warrant exceptional measures. The factors used to make these determinations were unknown. Given that the average Static-2002 scores were higher in the preselected high risk samples than the routine CSC samples (see Table 1), the decision-makers must have considered some factors already included in Static-2002 (e.g., prior offending). Given that the observed recidivism rates varied within scores, the decision-makers must have also considered valid factors not included in Static-2002. Previous research with Static-99, for example, has identified a number of risk factors that provide incremental validity beyond Static-99, including age (Barbaree & Blanchard, 2008) and professional evaluations of psychological constructs such as deviant sexual interests, intimacy deficits and personality disorders (Hanson et al., 2007; Knight & Thornton, 2007; Olver et al., 2007; Thornton, 2002).

Since the 1990s (when the vast majority of the CSC offenders in this study were incarcerated), CSC has adopted a proactive rehabilitation strategy toward all sexual offenders based on the risk, need, responsivity (RNR) model of correctional interventions (Williams, 1996). This model specifies that human service interventions should be targeted to the higher risk cases, that treatment should focus on the factors related to recidivism risk (criminogenic needs), and be adapted to the learning style of the offenders (Bonta & Andrews, 2007). Most sexual offenders during this period would have also received supervised release into the community involving human service interventions (e.g., employment, substance abuse counselling, maintenance groups).

The RNR model has received considerable empirical support for general offenders (Andrews & Bonta, 2006; Landenberger & Lipsey, 2005; Wilson, Bouffard, & MacKenzie, 2005). Further, there is evidence independent of CSC samples indicating that completion of modern sexual offender treatment programmes shows incremental prediction of sexual recidivism beyond Static-99 scores (McGrath, Cumming, Livingston, & Hoke, 2003). Readers should note, however, that not all the routine CSC offenders would have attended or completed sexual offender treatment, and approximately 20% of the "routine" CSC offenders would have been detained until Warrant Expiry.

Variation in absolute recidivism rates is not a problem in many settings. For decisions concerning allocation of scarce resources, such as prioritizing treatment and supervision, evaluators can use risk tools to determine relative risk. In this context, Static-2002 can be usefully considered an empirically validated method of summarizing the extent to which certain features of the offender's history make him more or less risky than other sex offenders. Statements about relative risk can be meaningfully expressed as percentile ranks or as relative risk compared to some normative population (Hanson, in press).

The finding that most of the datasets contained more than a trivial number of "impossible" scores raises the need to consider procedures to minimize such errors in applied use. Training is important (Bonta, Bogue, Crowley, & Motiuk, 2001; Flores, Lowenkamp, Holsinger, & Latessa, 2006), as is the conscientiousness of the evaluators (Hanson et al., 2007). Given that many of the errors involved simple mechanical operations (e.g., total scores), such errors could be eliminated by computerized coding forms, which automatically add the points and prohibit internal contradictions.

The current study found that Static-2002 was more accurate than Static-99 for the prediction of sexual, violent and general recidivism. Further research is needed concerning the extent to which Static-2002 has other useful properties; in particular, can the Static-2002 subscales be used as valid indicators of psychologically meaningful attributes (i.e., sexual deviance, general criminality)? If it is found that Static-2002 assesses specific constructs, a further question is whether knowledge of these constructs actually aids applied risk assessment. In high stakes evaluations, it is common for evaluators to use more than one risk tool (Jackson & Hess, 2007) and for different risk tools to produce different results (Barbaree, Langton, & Peacock, 2006). In principle, conceptually meaningful subscales allow for empirically defensible methods of resolving the divergent findings of the different risk scales (Doren, 2004b).

Recommendations for Applied Risk Assessment

1. The predictive accuracy of Static-2002 is sufficient that it can be used in applied risk assessments with sexual offenders.
2. For decisions concerning the relative risk of recidivism, risk can be indexed in terms of percentile ranks or relative risk (e.g., calculated through Cox regression coefficients; Hanson, Lloyd, Helmus, & Thornton, 2008).
3. For decisions concerning the absolute rate of sexual recidivism, we recommend that evaluators present both the low (routine CSC) and preselected high risk estimates.
4. Until further research is available, evaluators wishing to restrict their recidivism rate estimates to a narrow portion of the full range will need to resort to professional opinion. Such judgements should be based on the extent to which the offender and his circumstances more closely resemble the conditions of typical CSC offenders, or the diverse group of offenders examined in this study that were preselected to be high risk.
5. For the estimation of violent and general (any) recidivism rates, separate tables are required for sexual offenders with adult victims (rapists) and sexual offenders with child victims (child molesters). As a general guideline, child molesters have victims age 13 or less, and rapists have unrelated victims aged 16 or older. Further research is needed to determine how best to categorize offenders with young teen victims or those with a mixture of adult and child victims.

The fourth recommendation above requires evaluators to make judgements that are materially more complex than simply scoring an actuarial risk instrument. This complexity will no doubt provide fertile ground for forensic argument in high stakes evaluations. In general, judgements of this complexity are improved when structured by empirically based guidelines. Developing such guidelines, however, requires further research.

Acknowledgments The views expressed are those of the authors and not necessarily those of Public Safety Canada or the Wisconsin Department of Health Services. We would like to thank Howard Barbaree, Tony Beech, Susanne Bengtson, Jacques Bigras, Sasha Boer, Andy Haag, Leigh Harkins, Ray Knight, Calvin Langton and Jean Proulx for permission to use their data, and being patient with our ongoing questions.

APPENDIX

See appendix Tables 5, 6 and 7.

Table 5 Logistic regression summaries for the prediction of sexual, violent or any recidivism for Static-2002

Outcome	*k*	*n* recid/*N*	B_0 (*SE*)	B_1 (*SE*)	*r*	Goodness of fit χ^2 (*df*)	*p*
Sexual recidivism							
Overall 5 year	8	282/1,923	−3.497 (.1909)	.2917 (.0273)	−.9367	5.60 (7)	.587
Overall 10 year	5	242/1,132	−2.846 (.2045)	.2616 (.0299)	−.9294	10.38 (7)	.168
High risk 5 year	3	210/960	−2.636 (.2368)	.2138 (.0332)	−.9412	3.90 (6)	.691
CSC 5 year	3	53/734	−4.353 (.4188)	.3275 (.0628)	−.9372	2.75 (6)	.839
High risk 10 year	2	192/661	−2.064 (.2415)	.1879 (.0348)	−.9317	5.53 (6)	.478
CSC 10 year	2	29/342	−4.327 (.5890)	.3502 (.0872)	−.9403	5.85 (7)	.557
Rapist 10 year	5	46/231	−2.777 (.6938)	.2088 (.0990)	−.9708	7.08 (4)	.132
Child molester 10 year	5	53/348	−3.391 (.3828)	.2848 (.0519)	−.9110	3.91 (7)	.790
Sexual or violent recidivism							
Overall 5 year	7	415/1,732	−2.736 (.1573)	.2780 (.0239)	−.9271	15.30 (7)	.032
Overall 10 year	5	392/1,142	−2.267 (.1726)	.2839 (.0268)	−.9239	24.72 (7)	.001
Rapist 5 year	5	111/373	−2.750 (.4431)	.2978 (.0654)	−.9646	7.29 (5)	.200
Child molester 5 year	5	84/494	−3.031 (.2899)	.2621 (.0422)	−.9011	5.67 (7)	.579
High risk 5 year	2	248/769	−2.023 (.2240)	.2043 (.0323)	−.9352	8.04 (7)	.329
CSC 5 year	3	139/734	−3.089 (.2646)	.3128 (.0430)	−.9282	8.50 (6)	.204
High risk 10 year	2	283/671	−1.579 (.2172)	.2072 (.0326)	−.9283	14.15 (6)	.028
CSC 10 year	2	79/342	−3.076 (.3731)	.3574 (.0606)	−.9301	7.55 (7)	.374
Any recidivism							
Overall 5 year	7	668/1,735	−2.172 (.1352)	.3129 (.0221)	−.9203	18.81 (7)	.009
Overall 10 year	5	586/1,149	−1.665 (.1561)	.3162 (.0266)	−.9132	25.51 (7)	.001

The following formula can be used to estimate the recidivism rate ($\hat{p}$) for a specific Static-2002 score: $\hat{p} = \frac{e^{B0+B1\times\text{Score}}}{1+e^{B0+B1\times\text{Score}}}$

The standard error of the logit of $\hat{p}$ is $\sqrt{SE_{B0}^2 + (2 \times \text{Score} \times r \times SE_{B0} \times SE_{B1}) + \left[(\text{Score})^2 \times SE_{B1}^2\right]}$ (Fleiss et al., 2003, p. 295)

Table 6 Observed and estimated 5 year sexual recidivism rates: Complete sample

Score	Life table estimates		Fixed follow-up		Logistic regression estimates		
	Initial *n*	Recidivism rate (%)	Recidivists/total	Observed recidivism rate (%)	Predicted recidivism rate	95% C.I.	
0	67	1.5	1/30	3.3	2.9	2.0	4.2
1	195	2.8	2/106	1.9	3.9	2.8	5.3
2	310	4.3	8/176	4.5	5.1	4.0	6.7
3	355	4.3	10/193	5.2	6.8	5.5	8.4
4	435	8.1	25/247	10.1	8.9	7.5	10.5
5	448	13.5	41/282	14.5	11.5	10.0	13.2
6	387	12.8	36/271	13.3	14.8	13.2	16.6
7	286	17.1	36/206	17.5	18.9	17.0	21.0
8	273	23.8	49/200	24.5	23.8	21.1	26.7
9	137	30.6	33/99	33.3	29.5	25.8	33.5
10	94	36.8	27/90	38.6	35.9	30.9	41.3
11	34	26.8	8/31	25.8	42.8	36.3	49.6
12	12	50.6	6/11	54.5	50.1	42.1	58.0
Total	3,034	13.0	282/1,923	14.7			

Table 7 Static-2002 recidivism tables (Generated from logistic regression)

Static-2002 score	5 Year sexual recidivism (%)		10 Year sexual recidivism (%)	
	Routine CSC samples	Preselected high risk samples	Routine CSC samples	Preselected high risk samples
0	1.3	6.7	1.3	11.3
1	1.8	8.2	1.8	13.3
2	2.4	9.9	2.6	15.6
3	3.3	12.0	3.6	18.2
4	4.6	14.4	5.1	21.2
5	6.2	17.3	7.1	24.5
6	8.4	20.5	9.7	28.1
7	11.3	24.2	13.3	32.1
8	15.0	28.4	17.9	36.3
9	19.7	32.9	23.6	40.8
10	25.4	37.8	30.5	45.4
11[a]		43.0		50.1
12[a]		48.2		54.8
Total N[b]	734	960	342	661

[a] There were insufficient cases (<10) to produce reliable estimates for scores of 11 and 12 in the routine CSC samples. No offenders with a score of 13 or 14 were found in either sample

[b] N is the total sample size used in the logistic regression analysis to generate predicted recidivism values

REFERENCES

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

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Cincinnati, OH: NexisLexis.

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*, 84–94. doi:10.1207/s15327752jpa8701_07.

Barbaree, H. E., & Blanchard, R. (2008). Sexual deviance over the lifespan: Reductions in deviant sexual behaviour in the aging sex offender. In D. R. Laws & W. T. O'Donohue (Eds.), *Sexual deviance: Theory, assessment, and treatment* (2nd ed.). New York: Guilford.

Barbaree, H. E., Langton, C., & Peacock, E. J. (2006). Different actuarial risk measures produce different risk rankings for sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 18*, 423–440. doi:10.1177/107906320601800408.

Bengtson, S. (2008). Is newer better? A cross-validation of the Static-2002 and the Risk Matrix 2000 in a Danish sample of sexual offenders. *Psychology, Crime & Law, 14*, 85–106. doi:10.1080/10683160701483104.

Bigras, J. (2007). La prédiction de la récidive chez les délinquants sexuels [Prediction of recidivism among sex offenders]. *Dissertations Abstracts International, 68*(09). (UMI No. NR30941).

Boer, A. (2003). *Evaluating the Static-99 and Static-2002 risk scales using Canadian sexual offenders*. Unpublished master's thesis, University of Leicester, Leicester, United Kingdom.

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and rehabilitation (Corrections Research User Report No. 2007–06)*. Ottawa, ON: Public Safety Canada.

Bonta, J., Bogue, B., Crowley, M., & Motiuk, L. (2001). Implementing offender classification systems: Lessons learned. In G. A. Bernfeld, D. P. Farrington, & A. W. Leschied (Eds.), *Offender rehabilitation in practice* (pp. 227–245). Chichester, NY: Wiley.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668–1674. doi:10.1126/science.2648573.

Doren, D. M. (2004a). Stability of the interpretative risk percentages for the RRASOR and Static-99. *Sexual Abuse: A Journal of Research and Treatment, 16*, 25–36. doi:10.1177/107906320401600102.

Doren, D. M. (2004b). Toward a multidimensional model for sexual recidivism risk. *Journal of Interpersonal Violence, 19*, 835–856. doi:10.1177/0886260504266882.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). Hoboken, NJ: Wiley.

Flores, A. W., Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2006). Predicting outcome with the Level of Service Inventory-Revised: The importance of quality assurance. *Journal of Criminal Justice, 34*, 523–529. doi:10.1016/j.jcrimjus.2006.09.007.

Haag, A. M. (2005). Do psychological interventions impact on actuarial measures: An analysis of the predictive validity of the Static-99 and Static-2002 on a re-conviction measure of sexual recidivism. *Dissertation Abstracts International, 66*(08), 4531B. (UMI No. NR05662).

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. (2008). What statistics should we use to report predictive accuracy? *Crime Scene, 14*(1), 15–17. Retrieved September 15, 2008 from http://www.cpa.ca/sections/criminaljustice/publications/.

Hanson, R. K. (in press). The psychological assessment of risk for crime and violence. *Canadian Psychology/Psychologie canadienne*.

Hanson, R. K., Harris, A. J. R., Scott, T.-L., & Helmus, L. (2007). Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project (Corrections Research User Report No. 2007-05). Ottawa, ON: Public Safety Canada.

Hanson, R. K., Lloyd, C. D., Helmus, L., & Thornton, D. (2008). *Using multiple samples to estimate percentile ranks and relative risk ratios for actuarial risk tools: A Canadian example using Static-99 and Static-2002.* Unpublished manuscript.

Hanson, R. K., & Morton-Bourgon, K. E. (in press). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*.

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. doi:10.1023/A:1005482921333.

Hanson, R. K., & Thornton, D. (2003). Notes on the development of Static-2002. (Corrections Research User Report No. 2003-01). Ottawa: Department of the Solicitor General of Canada.

Harkins, L., & Beech, A. R. (2007). *Examining the effectiveness of sexual offender treatment using risk band analysis*. Manuscript submitted for publication.

Harris, A. J. R., Helmus, L., Hanson, R. K., & Thornton, D. (2008, October). *Are new norms needed for Static-99?* Paper presented at the 27th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Atlanta, GA.

Harris, A. J. R., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static-99 coding rules: Revised 2003*. Ottawa: Department of the Solicitor General of Canada.

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumiere, M. L., Boer, D., & Lang, C. (2003). A multi-site comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*, 413–425. doi:10.1037/1040-3590.15.3.413.

Hedges, L. V. (1994). Fixed effects 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.

Helmus, L. (2007). *A multisite comparison of the validity and utility of the Static-99 and Static-2002 for risk assessment with sexual offenders*. Unpublished B.A. (Honours) thesis, Carleton University, Ottawa, Ontario, Canada.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.

Humphreys, L. G., & Swets, J. A. (1991). Comparison of predictive validities measured with biserial correlations and ROCs of signal detection theory. *The Journal of Applied Psychology, 76*, 316–321. doi:10.1037/0021-9010.76.2.316.

Interstate Commission for Adult Offender Supervision. (2007). Sex offender assessment information survey (ICAOS Documents No. 4-2007). Lexington, KY: Author.

Jackson, R. L., & Hess, D. T. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sexual Abuse: A Journal of Research and Treatment, 19*, 409–448.

Knight, R. A., & Thornton, D. (2007). Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders (Document No. 217618). U.S. Department of Justice.

Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology, 1*, 451–476. doi:10.1007/s11292-005-3541-7.

Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007). Actuarial assessment of risk for reoffense among adult sex offenders: Evaluating the predictive accuracy of the Static-2002 and five other instruments. *Criminal Justice and Behavior, 34*, 37–59. doi:10.1177/0093854806291157.

McGrath, R. J., Cumming, G. F., & Burchard, B. L. (2003a). *Current practices and trends in sexual abuser management: The Safer Society 2002 Nationwide Survey*. Brandon, VT: Safer Society Foundation, Inc.

McGrath, R. J., Cumming, G., Livingston, J. A., & Hoke, S. E. (2003b). Outcome of a treatment program for adult sex offenders: From prison to community. *Journal of Interpersonal Violence, 18*, 3–17. doi:10.1177/0886260502238537.

Mossman, D. (2006). Another look at interpreting risk categories. *Sexual Abuse: A Journal of Research and Treatment, 18*, 41–63. doi:10.1177/107906320601800104.

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale—Sexual Offender Version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*(3), 318–329. doi:10.1037/1040-3590.19.3.318.

Phenix, A., Doren, D., Helmus, L., Hanson, R. K., & Thornton, D. (2008). *Coding rules for Static-2002*. Retrieved November 16, 2008 from http://www.static99.org/pdfdocs/static2002codingrules.pdf.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737–748. doi:10.1037/0022-006X.63.5.737.

Soothill, K. L., & Gibbens, T. C. N. (1978). Recidivism of sexual offenders: A re-appraisal. *The British Journal of Criminology, 18*, 267–276.

Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research and Treatment, 14*, 139–153. doi:10.1177/107906320201400205.

Williams, S. (1996). *Standards and guidelines for the provision of services to sex offenders*. Ottawa: Correctional Service of Canada. Retrieved September 16, 2008 from http://198.103.98.138/text/pblct/so/standards/stande-eng.shtml.

Wilson, D. B., Bouffard, L. A., & Mackenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders. *Criminal Justice and Behavior, 32*, 172–204. doi:10.1177/0093854804272889.