Predictive Validity of the Static–99 and Static–2002 for Sex Offenders on Community Supervision

Leslie M. D. Helmus[1] *& R. Karl Hanson*[2]
[1]Corrections Research, Public Safety Canada; [2]Carleton University Ottawa, Canada

Abstract

The Static–99 is the most commonly used actuarial tool for sexual offenders. Although it has shown acceptable predictive accuracy in a large number of studies, all these studies involved researchers scoring the instrument retrospectively. Consequently, it is unclear whether similar results would be obtained when used in routine practice. The authors of the Static–99 have proposed a new scale, the Static–2002, but there has been insufficient research to determine whether it is an improvement over the Static–99. This study examined the predictive accuracy of the Static–99 in a prospective study of 706 Canadian sexual offenders on community supervision. All assessments were conducted by the probation and parole officers responsible for supervising the cases. The Static–99 was compared with the Static–2002, which was scored retrospectively from criminal history records. After an average 3 year follow–up, the Static–99 and Static–2002 were equally accurate in predicting sexual recidivism (ROC of .76 for both). The Static–2002, however, was better than the Static–99 at predicting violent and general recidivism. There were no significant differences in the accuracy of the measures for rapists, child molesters, or non–contact offenders. Overall, the Static–99 and Static–2002 are both reliable and valid measures of recidivism risk for sexual offenders.

Key words: Static–99, Static–2002, recidivism, risk assessment, sexual offenders

Risk assessment has become embedded in our criminal justice system with important consequences for public safety – and the offender. Although there are various approaches to risk assessment (Hanson, 1998), actuarial risk tools have attracted increased attention during the past ten years (Archer, Buffington–Vollum, Stredny, & Handel, 2006).

Since its introduction in 1999, the Static–99 (Hanson & Thornton, 2000) has quickly become the most widely used and widely researched risk assessment instrument for sexual offenders (Archer et al., 2006). A 2002 nation–wide survey of sex offender treatment providers in the United States found that 54% of community programs used the Static–99 (McGrath, Cumming, & Burchard, 2003). This rapid adoption is particularly impressive given that the survey was conducted only two years after the Static–99 was published. Currently it is the most widely used risk tool for sexual offenders in the United States (Interstate Commission for Adult Offender Supervision, 2007) and Canada, and is used in countries as diverse as Israel, Singapore, and Taiwan.

The Static–99 is also the most researched of all risk assessment tools for sex offenders. Hanson and Morton–Bourgon (2007) found 42 Static–99 replication studies, with the average predictive accuracy similar to that found in the original developmental samples. Reviews, however, have found significant variability across studies. Some of this variability may be due to differences in the

offenders. The predictive accuracy of the Static–99 typically is similar for child molesters and rapists (Bartosh, Garby, & Lewis, 2003; Ducro & Pham, 2006). Bartosh et al., however, found low predictive accuracy (ROC = .39) among non–contact offenders, although the sample size was small (n = 17). Further research is needed to explain the observed variation in findings across studies.

The reasons for the popularity of the Static–99 are unknown, but are probably related to the ease with which it can be scored from commonly available criminal history information. Although more complex systems may improve upon the predictive accuracy of the Static–99 for estimating the likelihood of sexual recidivism (Hanson, Harris, Scott, & Helmus, 2007; Thornton, 2002), and violent recidivism (G. T. Harris, Rice, Quinsey, Lalumiere, Boer, & Lang, 2003), the test appeals to test–users seeking a cost–effective evaluation applicable to a wide range of sexual offenders.

Despite its success, it is far from ideal. The predictive accuracy is only moderate and the scores have no intrinsic meanings. The items were selected based on their ease of administration and association with recidivism. Evaluators using more than one actuarial scale will often find divergent risk rankings (Barbaree, Langton, & Peacock, 2006), and these conflicting results can only be explained by an understanding of what the scales are measuring. For example, some scales may be heavily weighted with items associated with sexual deviancy and other scales may be weighted with items associated with general criminality (Doren, 2002).

In order to address the weaknesses in the Static–99, Hanson and Thornton (2003) proposed a new scale entitled the Static–2002. It was hoped that the Static–2002 would retain the advantages of the Static–99, namely that it would be applicable to numerous samples and easily scored (Hanson & Thornton, 2003), and would be better in four ways:

1. Increased coherence and conceptual clarity. The Static–99 is a second–generation (see Bonta, 1996), "atheoretical" risk assessment instrument. The Static–2002 contains items that were intended to measure conceptually meaningful constructs, such as sexual deviancy, general criminality, and persistence.
2. Improved consistency of the scoring criteria. Because the Static–99 items came from two separate instruments (the RRASOR and the SACJ–Min), its coding rules are inconsistent. Depending on the item, convictions, charges, or sentencing occasions are counted. The Static–2002 was designed to reduce these inconsistencies and thereby facilitate training and increase inter–rater reliability.
3. Reduced counter–intuitive scorings. With the Static–99, there are rare cases where it is possible for an offender to be scored, commit a new sexual offence, and receive a *lower* score when scored again. This can occur if the previous offence included non–sexual violence, but the recidivism incident (the new index offence) does not. This counter–intuitive result is related to the Static–99's origins in "dustbowl" empiricism (Andrews & Bonta, 2003, p. 238), whereby items were not considered in any depth beyond their predictive accuracy. The Static–2002 eliminates this coding possibility.
4. Increased predictive accuracy. The Static–2002 contains more items than the Static–99 and definitions were revised based on evidence. It was hoped that these modifications might increase the predictive accuracy of the measure.

The Static–2002 has 13 items (see Table 1). Some items are the same as in the Static–99 (either with the same or modified coding rules) and some are new items. Notably, two Static–99 items were not included in the Static–2002. The item regarding intimate relationships (never lived with a lover for at least two years) was removed because it can be difficult to score in adversarial contexts (Hanson & Thornton, 2003). Additionally, the item for non–sexual violence conviction during the index offence was deleted because it was the source of the counter–intuitive scorings discussed

earlier (Hanson & Thornton, 2003). Similar to the Static–99, items were included based on their empirical relationship with recidivism; however, increased attention was paid to scoring consistency (e.g., convictions versus sentencing occasions). Also, to increase coherency and clarity of the constructs being measured, the items in the Static–2002 are grouped into five domains: age, persistence of sex offending, deviant sexual interests, relationship to victims, and general criminality.

Table 1: Static–2002 Items

| Static–2002 Domain and Item | Included in Static–99 | | New |
	Same Coding	Modified Coding	
Age			
Age	–	Yes	–
Persistence of sexual offending			
Prior sex offences	–	Yes	–
Juvenile arrest for sex offence	–	–	Yes
High rate of sex offending	–	–	Yes
Sexual deviance			
Non–contact convictions	Yes	–	–
Male victims	Yes	–	–
2+ victims, at least one unrelated	–	–	Yes
Relationship to victim			
Unrelated victims	Yes	–	–
Stranger victims	Yes	–	–
General criminality			
Prior arrest/sentencing occasions	–	Yes	–
Breach of conditional release	–	–	Yes
4 years free prior to index	–	–	Yes
Prior non–sexual violence	–	Yes	–

In the development study, Hanson and Thornton (2003) compared the predictive accuracy of the Static–99 and Static–2002 by combining eight samples. This should be considered a rough comparison due to missing information: none of the samples had information to code breach of conditional release and time free prior to index. Despite this missing data, the Static–2002

performed slightly better than the Static–99 for predicting sexual recidivism (ROC values of .71 and .70, respectively). Similar to Hanson and Morton–Bourgon's later (2007) meta–analysis, the Static–99 showed significant variability across samples, whereas the Static–2002 was more consistent (variability was not significant). For the prediction of violent recidivism, Hanson and Thornton (2003) found that the Static–2002 significantly outperformed the Static-99 (ROC values of .71 and .69, respectively).

In Hanson and Morton–Bourgon's (2007) meta–analysis, there are only five replications of the Static–2002. The average effect size (d) for the Static–2002 was .78 for the prediction of sexual recidivism, compared to .70 for the 42 replications of the Static–99. For the prediction of violent recidivism, the Static–2002 showed a stronger advantage over the Static–99 (d = .71 versus .58). This is likely due to more items assessing general criminality in the Static–2002. There were insufficient studies of the Static–2002 (less than three) to evaluate its ability to predict any recidivism.

The purpose of this study was to directly compare the Static–99 and the Static–2002 in a large sample of Canadian sexual offenders on community supervision. An important feature of this prospective study was that the Static–99 was scored by the officers responsible for the supervision of the cases. In previous studies, researchers scored the Static–99 retrospectively from existing file information. An additional goal was to examine the validity of the measures across offender subgroups, particularly given the limited research suggesting that the Static–99 may not be valid with non–contact sex offenders (Bartosh et al., 2003). This study was part of the larger Dynamic Supervision Project (Hanson et al., 2007). Although the purpose of the Dynamic Supervision Project was to validate the Stable–2000 and Acute–2000 as part of a dynamic risk assessment protocol for sex offenders, this project also afforded an opportunity to validate the Static–99 and Static–2002 on a diverse sample.

Method

Subjects

All offenders were adults starting a period of community supervision (probation or parole) for a recent sexual offence. A sexual offence was defined as an offence with a sexual motivation involving a non–consenting person or persons unable to provide consent (Category "A" offences in the Static–99 coding rules; A. J. R. Harris, Phenix, Hanson, & Thornton, 2003). The vast majority of the offenders were currently serving sentences for sexual offences, although a small number had a conviction for a sexual offence within the previous two years but were currently being supervised for a non–sexual offence. In all cases, the supervising officers considered the offenders' primary problem to be sexual offending and were supervising them as such. Offenders were excluded if they had been in the community for a period of six months prior to initial assessment, had successfully appealed their conviction, were serving sentences for crimes committed prior to the age of 18, or had only been convicted of sexual offences involving consenting adults (e.g., prostitution).

The offenders in the larger study came from 16 different jurisdictions: all Canadian provinces and territories, the Atlantic Region of the Correctional Service of Canada, and the states of Alaska and Iowa. In total, 156 probation officers submitted assessment information on 997 offenders. For the current study, the cases from Alaska and Iowa (n = 201) were deleted because missing criminal history data precluded coding of the Static–2002. Additionally, 90 cases were deleted because the Static–99 was not submitted, the Static–2002 could not be coded due to missing Canadian criminal history data, or the offender had not yet been released from custody. Four female offenders were

also deleted because risk assessment procedures developed for male sex offenders are unlikely to be accurate when applied to female sex offenders (Cortoni & Hanson, 2005). This left 702 offenders for the current study. Descriptive information on the sample is presented in Table 2. On average, the offenders were 42 years old and serving their first sentence for a sexual offence (66%). Half the sample had child victims, and approximately 8% had only non–contact offences. Most of the sexual offences (78%) involved physical contact but not overt physical injury. Approximately 15% of the offenders self–identified as being of Aboriginal heritage, 11% had been hospitalized overnight for a psychiatric condition, and 5% had previously been diagnosed as developmentally delayed.

Table 2: Descriptive statistics (N = 702)

	n (%)	n missing
Age at release M (SD)	41.6 (13.2)	1
Offender Type		28
Non–contact	53 (7.9%)	
Child molester	337 (50.0%)	
Rapist	244 (36.2%)	
Mixed	40 (5.9%)	
Most serious victim injury		1
Non–contact offences only	58 (8.3%)	
Physical contact	546 (77.9%)	
Victim injury	90 (12.8%)	
Life threatening injury	7 (1.0%)	
Any prior sexual offences	236 (33.6%)	0
Developmentally delayed	33 (4.8%)	16
Major mental disorder	78 (11.3%)	14
Aboriginal	106 (15.2%)	8
Static–99 M (SD)	2.86 (2.0)	0
Static–2002 M (SD)	4.06 (2.3)	0
Average Follow–up: Years (SD)	3.2 (1.2)	0
Recidivism Rate		
Sexual (*N* = 702)	57 (8.1%)	0
Rapists (*n* = 244)	25 (10.2%)	0

Child molesters (*n* = 337)	14 (4.2%)	0
Non–contact (*n* = 53)	6 (11.3%)	0
Violent (*N* = 702)	115 (16.4%)	0
Rapists (*n* = 244)	56 (23.0%)	0
Child molesters (*n* = 337)	33 (9.8%)	0
Non–contact (*n* = 53)	9 (17.0%)	0
Any (*N* = 702)	196 (27.9%)	0
Rapists (*n* = 244)	91 (37.3%)	0
Child molesters (*n* = 337)	68 (20.2%)	0
Non–contact (*n* = 53)	14 (26.4%)	0

Measures

Static–99 (Hanson & Thornton, 2000)

The Static–99 is an actuarial risk tool designed to predict the probability of recidivism among adult male sexual offenders. It has 10 items: young, never lived with a lover for 2+ years, prior sex offences, four or more prior sentencing dates, any convictions for a non–contact offence, index non–sexual violence, prior non–sexual violence, any unrelated victims, any stranger victims, and any male victims. Items are coded as either a 0 or a 1, except for prior sexual offences, which is scored as 0, 1, 2, or 3. Total scores (obtained by summing all the items) can range from 0–12. Based on an offender's total score, they are placed in one of four risk categories: low (0–1), moderate–low (2–3), moderate–high (4–5), and high (6+). Previous research has demonstrated high rater reliability and moderate predictive accuracy (see A. J. R. Harris et al., 2003).

Static–2002 (Hanson & Thornton, 2003)

The Static–2002 was designed for the same purpose as the Static–99, namely, assessing the recidivism risk of adult male sexual offenders based on commonly available criminal history information. The 13 items are displayed in Table 1. Total scores can range from 0–14. Based on an offender's total score, they are placed in one of five risk categories: low (0 – 2), low–moderate (3, 4), moderate (5, 6), moderate–high (7, 8), and high (9+; Helmus, 2007). Previous research has demonstrated high rater reliability (Langton, Barbaree, Hansen, Harkins, & Peacock, 2007) and moderate to high predictive accuracy (Hanson & Morton–Bourgon, 2007).

Procedure

Data Collection

Static–99

Static–99 risk assessments were coded prospectively by probation and parole officers participating in the Dynamic Supervision Project. Data were collected as part of their routine supervision practices. The assessment data were considered administrative records controlled by the specific

jurisdictions and did not require the consent of offenders to collect. Formal agreements were developed with the participating jurisdictions to share the data with the researchers at Public Safety Canada (then Solicitor General Canada) for the common purpose of program evaluation. In all cases, the data remained the property of the specific jurisdictions, and the researchers at Public Safety Canada were in the role of data managers. For the province of Quebec, however, the consent of the offenders was required before the data could be shared with a federal government department. Consequently, the offenders from Quebec all signed consent forms allowing their assessment data to be used for this study.

Following training, participating officers were requested to submit information on consecutive, new cases until a sufficient sample size had been collected (estimated to be three years). File review conducted for the purpose of the reliability training suggested that the selection of cases was not always consecutive. The reasons for scoring only some new cases are unknown, but appeared to be related to the degree of local administrative support for the project and the competing time demands placed on the officers. The officers submitting data to the project were volunteers. Some jurisdictions required officers to attend the training, and other jurisdictions adopted all or some of the measures as standard practice; however, the decision to submit data to the research project was at the initiative of the individual officers.

The requested data collection protocol involved submitting the Static–99 within the first month of supervision. In practice, these assessments took approximately 30 minutes to complete. There was no missing information in the assessments. In addition to the Static–99 items, the officers submitted basic demographic information for the offender as well as the ages and pre–existing relationships for all known victims of sexual offences.

Static–2002

The Static–2002 was retrospectively coded in 2005 by the first author and a M.A. student, using the original offence history variables coded by the probation officers, and Canadian Police Information Centre (CPIC) records maintained by the Royal Canadian Mounted Police (RCMP). CPIC records contain basic criminal history information; namely, date of conviction, offence title (according to the Canadian Criminal Code), the sentence, and the police jurisdiction that reported the incident. Information on charges that were stayed or for which the offender received an acquittal are inconsistently recorded.

Training

All officers submitting data to the project were required to attend a two day training session which covered the Static–99, the Stable–2000, and the Acute–2000. Most of the training sessions were conducted by Karl Hanson and Andrew Harris, although other trainers were used in some jurisdictions. In rare cases, officers submitted data who had been trained by apprenticing with other local officers. The training primarily involved descriptions of the scoring criteria and structured exercises (both oral and written). The final written scoring exercise was collected for reliability purposes (see below). The training did not have any formal pass/fail criteria, but officers who were having obvious difficulty were encouraged to review their work with another officer before submitting it to the project.

Data were submitted via fax or secure website. In the initial registration, the offenders' identifying information was linked to a unique identifier known only to the research team and the supervising officer. In order to protect confidentiality, all subsequent data submissions included only the unique identifier.

Reliability

Static–99

Reliability for the Static–99 assessments completed by the supervising officers was examined in two ways. The first method involved comparing the officers' responses in the final training exercise to the responses of Karl Hanson and Andrew Harris. Training exercises were analyzed for 213 officers trained by Hanson and Harris and 45 officers trained in Ontario by Susan Cox, Mark Stehlin, and Donna–Lee Rabey–McKay. Approximately one third of the officers completed one of three different exercises. The reliability is reported in terms of "percent correct" because there was artificial variability in the training exercises and an answer key that was presumed to be correct.

For officers trained by the principal investigators, between 91% and 95% (depending on the exercise) were within one point of the correct answer for the total Static–99 score. For the Ontario officers trained by other trainers, between 88% and 91% were within one point of the correct answer for the total Static–99 score. These results indicate that it is possible to train trainers who can train as effectively as the original test developers.

The second method of checking reliability involved file reviews of 88 cases registered with the project. The review cases were selected randomly from the settings providing the largest number of cases as well as additional locations that fit into the travel schedule of the expert raters[1]. Both criteria resulted in over–sampling major urban settings. The reviewers' task was to identify the best scoring given the file information available. The reliability calculated from this approach would tend to overestimate rater agreement because a) the second raters were not blind to the previous ratings, b) both ratings were based on the same file information coded by the original rater, and c) the second rater was able to question the original rater about information that was missing or ambiguous. Consequently, these reliability reviews should be considered a test of the extent to which the officers understood the coding rules rather than a test of the concordance between fully independent assessments. The agreement was high between the original ratings and the consensus ratings developed through file reviews. The intraclass correlation (ICC) for the Static–99 was .91 ($k = 88$).

Static–2002

Inter–rater reliability for the Static–2002 was calculated for 25 cases. The Intraclass Correlation Coefficient (ICC) was .98. This is exceptionally high and should not be considered representative of the typical circumstances where the Static–2002 would be coded. In this study, all victim information was already identified and coded by the supervising officer who submitted the Static–99. Additionally, because the CPIC records were used, criminal history was coded based on the Criminal Code offence, and not on the circumstances of the incident. These sources simplified the task and contributed to the artificially high reliability.

Recidivism

Information concerning new offences was gathered through reviews of provincial and national (Canadian) criminal history records, as well as from supervising officers and local police jurisdictions. CPIC records were received in August, 2005, and June, 2006. Provincial records were received from the following jurisdictions: British Columbia (January, 2006), Manitoba (April, 2005), and Ontario (December, 2005). The Offender Management System of the Correctional Service of Canada (CSC) was checked in May, 2005, for recidivism information of the CSC offenders registered in the project.

Once a recidivism event was identified, we then obtained the date on which the offence occurred, and a brief description of the offence behaviour in order to classify it into one of three categories (see below). Offence information was provided by the supervising officers, provincial or state correctional systems, and through direct contact with the police jurisdictions responsible for the original charges. In some cases, the police provided information about new offences that had yet to appear on other records. Consequently, the last known recidivism event was in February, 2007 – the date of our last follow–up with a police jurisdiction.

The follow–up period was calculated from the date that the first assessment information was collected to the date of the last recidivism information received. For the few cases that did not appear on any official record, the follow–up end date was set one month after the last assessment information was received. The offender start dates ranged from January 18, 2001 to October 19, 2005, with a mean follow–up time of 3.2 years (M SD = 1.2, range of .03 to 5.4 years). For the purpose of survival analyses, the start date was the date of first assessment or the date of release into the community, whichever was latest. The survival end date was the earliest of the following events: sexual recidivism, death, deportation, end of follow–up, or incarceration for a period of time that included the follow–up end date.

Three types of recidivism were used in this study. The first category was "sexual crime recidivism", which included all crimes with a sexual motivation, whether or not the name of the offence was explicitly sexual. This included contact and non–contact offences, as well as sexual offences involving consenting adults (e.g., prostitution, public sex). "Violent recidivism" was defined as all crimes that involved direct confrontation with the victim, and included sexual crime recidivism. Given that some of the sexual offences were not violent (e.g., prostitution), this category could also be called "violent or sexual recidivism". The final category, "any recidivism", included all crimes (sexual, violent, and non–violent) as well as all breaches.

Criminal recidivism was considered to have occurred if the agency reporting the information believed that the offence occurred. For example, we counted as a sexual recidivist a man who self–reported to his therapist that he had exposed himself even though no official sanctions were initiated at that time. For breaches, however, an official record of parole revocation or a new conviction for violation of conditional release was required. Given that criminal history records were the major source of recidivism information, the vast majority of recidivism events were linked to an officially recorded charge or conviction.

Plan of Analysis

Effect sizes were coded as Receiver Operating Characteristic (ROC) areas, which is one of the most commonly used and commonly recommended effect sizes for risk prediction (Rice & Harris, 2005). ROC areas are typically preferred to correlations or Cohen's d because they are not affected by the base rate of the event (Rice & Harris, 2005). ROC curves plot the false positive rate by the correct positive rate for each possible cut–off score in the prediction scheme, which creates a curve (Swets, Dawes, & Monahan, 2000). ROC areas refer to the Area Under the Curve (AUC) and can vary between 0 and 1. If an ROC value is .5, this is the level of prediction that would be expected by chance. An ROC value less than .5 indicates prediction at less than chance. ROC values between .5 and 1 indicate prediction exceeding chance levels, with numbers closer to 1 showing stronger prediction. Because .5 indicates chance level, a confidence interval that does not include .5 demonstrates predictive accuracy significantly greater (or less) than chance. Another way to interpret the ROC area is that it represents the probability that a randomly selected recidivist will have a higher risk score than a randomly selected non–recidivist (Rice & Harris, 2005). Also worth

noting is that ROC values of .56, .64, and .71 correspond to the small (.2), moderate (.5), and large (.8) effect size conventions of Cohen's *d* (Rice & Harris, 2005).

To test whether the Static–99 and Static–2002 differed in their level of predictive accuracy, Hanley and McNeil's (1983) test of correlated ROC areas was used. To examine differences between offender subgroups, the Q statistic was used (Hanson & Broom, 2005).

Results

Table 3 summarizes the predictive accuracy (ROC areas) for the Static–99 and Static–2002 for sexual, violent, and any recidivism. All confidence intervals do not include .5, indicating that the Static–99 and Static–2002 significantly predicted sexual, violent, and any recidivism. The predictive accuracy for sexual recidivism was virtually the same for both measures, with the Static–99 showing a slight, non–significant advantage over the Static–2002 (.764 versus .761, respectively. Difference = −.003, 95% C.I. of −.037 to .031). For the prediction of violent recidivism, the Static–2002 had greater predictive accuracy than the Static–99 (.764 versus .730, respectively). The difference between the two measures was .034, with a confidence interval between .008 and .060, which was statistically significant. For the prediction of any recidivism, the Static–2002 had greater predictive accuracy than the Static–99 (.759 versus .718, respectively). The difference between the two measures was .041, with a confidence interval between .019 and .063, which was statistically significant. Overall, although both measures predict sexual recidivism with the same accuracy, the Static–99's accuracy decreased as it predicted violent and any recidivism, whereas the Static–2002's accuracy remained at approximately the same level across all three outcomes.

Table 3: ROC Areas for the Static–99 and Static–2002

		Sexual Recidivism		Violent Recidivism		Any Recidivism	
Sample	*n*	ROC	95% C. I.	ROC	95% C. I.	ROC	95% C. I.
Static–99	702	.764	.699–.829	.730	.680–.780	.718	.676–.760
Static–2002	702	.761	.698–.823	.764	.720–.807	.759	.721–.797

Table 4 presents the ROC values for the Static–99 and Static–2002 across offender subgroups of rapists, child molesters, and non–contact sex offenders. Across all three outcomes, both the Static–99 and Static–2002 showed moderate to large effect sizes with child molesters, rapists, and non–contact offenders. All effect sizes were significant except for the Static–2002's prediction of violent and any recidivism for non–contact offenders. For both the Static–99 and the Static–2002, there were no significant differences in predictive accuracy across the three offender subgroups. When comparing the predictive accuracy of the two measures within subgroups, four significant differences emerged. The Static–2002 showed greater predictive accuracy than the Static–99 in the prediction of violent recidivism for rapists, violent recidivism for child molesters, any recidivism for rapists, and any recidivism for child molesters. However, when Bonferroni's correction for multiple comparisons was applied, the only finding that remained significant was that the Static–2002 predicted any recidivism in child molesters with greater accuracy than the Static–99 (difference =

.056, 99.4% C.I. of .002 to .110).

Table 4: ROC Areas for the Static–99 and Static–2002 across offender subgroups

Sample	n	Sexual Recidivism			Violent Recidivism			Any Recidivism		
		ROC	95%	C. I.	ROC	95%	C. I.	ROC	95%	C. I.
Rapists										
Static–99	244	.718	.610	.827	.664	.580	.749	.686	.617	.755
Static–2002	244	.703	.584	.821	.720	.642	.798	.732	.668	.796
Child Molesters										
Static–99	337	.758	.611	.904	.747	.663	.831	.710	.641	.779
Static–2002	337	.809	.697	.921	.798	.734	.863	.766	.705	.828
Non–Contact										
Static–99	53	.730	.558	.903	.667	.508	.825	.724	.583	.865
Static–2002	53	.652	.469	.836	.635	.480	.790	.700	.553	.846

Discussion

The Static–99 and Static–2002 showed the same level of accuracy in predicting sexual recidivism, and the Static–2002 was more accurate than the Static–99 at predicting violent and general recidivism. These findings are largely consistent with a recent meta–analytic review of eight published and unpublished Static–2002 studies (including the data reported in this study; Helmus, 2007). Helmus found that the Static–2002 was more accurate than the Static–99 at predicting violent and general recidivism, but she also found that the Static–2002 was significantly better than the Static–99 at predicting sexual recidivism (ROC values of .71 and .69, respectively). The difference in the predictive accuracy was small, however, and the variability for the Static–2002 across studies was no more than would be expected by chance. Consequently, a reasonable conclusion is that the Static–2002 is as good as the Static–99 at predicting sexual recidivism and better at predicting non–sexual recidivism. The Static–2002's stronger predictive accuracy for violent and any recidivism is expected given that the Static–2002 contains more items related to general criminality than does the Static–99.

These current results also suggested that both measures are equally valid for the major subgroups. In contrast to Bartosh et al. (2003), the Static–99 significantly predicted all three types of recidivism with non–contact offenders in the current study. Bartosh et al. found a surprisingly low ROC value of .39 for the prediction of sexual recidivism with a small sample of 17 non–contact offenders. The

current study found a large effect size (ROC = .73) for the Static–99's prediction of sexual recidivism with non–contact offenders. To compare the seemingly discrepant findings from the current study and Bartosh et al.'s study, a cumulative meta–analysis was performed according to the recommendations of Hanson and Broom (2005). The cumulative average ROC effect size (weighted by the inverse of the variance) was .631 for the two studies, with a 95% confidence interval between .49 and .78. The Q statistic was significant (Q = 4.310, p < .05), indicating that the two findings were significantly different from each other. It is likely that either differences in the subjects or in the procedures resulted in the diverse findings. Differences in the subjects are plausible because the offenders in Bartosh et al.'s study were released from incarceration after an average sentence of 66 months, whereas the majority of participants in the Dynamic Supervision Project were on community supervision without a period of incarceration, and less than 2% had been sentenced to more than 24 months of incarceration. Given that there are only two studies, each with a small sample size and potentially different populations, it is too soon to make strong conclusions about the validity of the Static–99 with non–contact offenders. Future research should shed light on whether the true accuracy is closer to the .39 estimate found by Bartosh et al. or the .73 estimate found in the current study.

In the current study, the predictive accuracy of the Static–2002 with non–contact offenders was lower than it was for the Static–99. The lack of statistical significance, however, was likely due to the small sample size. The effect sizes were still moderate to large and the results across subgroups were not significantly different or different from the Static–99 findings. It is necessary to conduct further research with larger samples of non–contact offenders to further examine both measures with this population.
The high reliability and the moderate to large predictive accuracy for the Static–99 demonstrated that the measure can be used reliably and accurately in applied settings by non–researchers. To our knowledge, this Static–99 replication is the only one that started with the routine assessments conducted by probation and parole officers supervising their cases. Prospective predictive accuracy is difficult to demonstrate because the assessment itself may trigger intervention strategies affecting the outcome. In this study, however, the interventions received by the offenders did not substantially change their relative risk rankings. Experience with the jurisdictions involved in this study would suggest that few, if any, of the high risk offenders would have received interventions that would substantially change their risk levels.

There are numerous strengths to this study. Firstly, comparing the same offenders on both instruments offers a more direct comparison than examining one measure with one sample and the other measure with a separate sample, which is the case with meta–analyses. Another strength is the large and diverse sample which enhances the generalizability of the findings. An additional advantage of this study is the comprehensive information for the Static–99 and Static–2002 items. In the Static–2002 development study, none of the samples had sufficient information to code two of the items (breach of conditional release, and time free prior to index), and many samples had additional missing information. In that study, cases were only included if the Static–99 had three missing items or less, and the Static–2002 had five missing items or less. This is still a substantial amount of missing information. The present study had no missing information, which increases the confidence in the obtained accuracy.

Limitations of this study were that the Static–2002 was coded retrospectively, by different raters than the officers who submitted the Static–99, and with less information. Despite these limitations, this study provided a rigorous comparison of the Static–99 and Static–2002. Future research is necessary to examine how other factors (e.g., exposure to treatment, jurisdiction) can influence predictive accuracy. Also, future replications of the Static–2002 on specialized samples (e.g., developmentally delayed, internet sex offenders, juveniles) can build on current knowledge. The

results of this study indicate that the Static–2002 predicts sexual recidivism as well as the Static–99, and predicts violent and any recidivism with greater accuracy than the Static–99. These results suggest that the Static–2002 could replace the Static–99 in applied risk assessments. Before the Static–2002 is widely implemented, however, an official user's manual and training protocols are needed.

References

1. Andrews, D. A., & Bonta, J. (2003). The psychology of criminal conduct. Cincinnati, OH: Anderson Publishing.
2. 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.
3. Barbaree, H. E., Langton, C. M., & Peacock, E. J. (2006). Different actuarial riskmeasures produce different risk rankings for sexual offenders. Sexual Abuse: A Journal of Research and Treatment, 18, 423–440.
4. Bartosh, D. L., Garby, T., & Lewis, D. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. International Journal of Offender Therapy & Comparative Criminology, 47, 422–438.
5. Bonta, J. (1996). Risk–needs assessment and treatment. In A. T. Harland (Ed.), Choosing correctional options that work: Defining the demand and evaluating the supply (pp. 18–32). Thousand Oaks, California: Sage Publications.
6. Cortoni, F., & Hanson, R. K. (2005). A review of the recidivism rates of adult female sexual offenders (Research Report No. R–169). Ottawa: Correctional Service of Canada.
7. Doren, D. M. (2004). Toward a multidimensional model for sexual recidivism risk. Journal of Interpersonal Violence, 19, 835–856.
8. Ducro, C., & Pham, T. (2006). Evaluation of the SORAG and the Static–99 on Belgian sex offenders committed to a forensic facility. Sexual Abuse: A Journal of Research and Treatment, 18, 15–26.
9. 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.
10. Hanson, R. K. (1998). What do we know about sexual offender risk assessment? Psychology, Public Policy and Law, 4, 50–72.
11. Hanson, R. K., & Broom, I. (2005). The utility of cumulative meta–analysis: Application to programs for reducing sexual violence. Sexual Abuse: A Journal of Research and Treatment, 17, 357–373.
12. 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 (User Report 2007–05). Ottawa, ON: Public Safety Canada.
13. Hanson, R. K., & Morton–Bourgon, K. (2007). The accuracy of recidivism risk assessments for sexual offenders: A meta–analysis (User Report 2007–01). Ottawa, ON: Public Safety and Emergency Preparedness Canada.
14. Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. Law and Human Behavior, 24(1), 119–136.
15. Hanson, R. K., & Thornton, D. (2003). Notes on the development of the Static–2002 (User Report 2003–01). Ottawa, ON: Solicitor General Canada.
16. Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. (2003). Static–99 coding rules: Revised 2003. Ottawa, ON: Solicitor General Canada.

17. 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.
18. Helmus, L. (2007). A multi–site comparison of the validity and utility of the Static–99 and Static–2002 for risk assessment with sexual offenders. Unpublished B.A. thesis, Carleton University, Ottawa, Ontario, Canada.
19. Interstate Commission for Adult Offender Supervision (2007, April). SO assessment information survey 4–2007. Retrieved July 20, 2007, from http://www.interstatecompact.org/resources/surveys/default.shtml
20. Langton, C. M., Barbaree, H. E., Hansen, K. T., Harkins, L., & Peacock, E. J. (2007). Reliability and validity of the Static–2002 among adult sexual offenders with reference to treatment status. Criminal Justice and Behavior, 34, 616–640.
21. McGrath, R. J., Cumming, G. F., & Burchard, B. L. (2003). Current practices and trends in sexual abuser management: The Safer Society 2002 nationwide survey. Brandon, VT: Safer Society Press.
22. Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow–up studies: ROC area, Cohen's d, and r. Law and Human Behavior, 29(5), 615–620.
23. Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. Sexual Abuse: A Journal of Research and Treatment, 14, 139–153.

Author address

Leslie M. D. Helmus
Corrections Research, Public Safety Canada
340 Laurier Ave. West, 10th floor
Ottawa, ON, Canada, K1A 0P8
Phone: 613–998–0312
Fax: 613–990–8295
E–Mail: leslie.helmus@ps.gc.ca

Psychology Department, Carleton University
1125 Colonel By Drive
Ottawa, ON, Canada, K1S 5B6