Applicability of the New Static-99 Experience Tables in Sexually Violent Predator Risk Assessments

Brian R. Abbott
Private Practice, San Jose, CA

Abstract

The Static-99 is an actuarial risk assessment tool often used in the evaluation of sexually violent predators. The risk estimates from the original Static-99 developmental sample (experience tables ; Hanson & Thornton, 2000) have been criticized as being inaccurate when applied to contemporary groups of sexual offenders including sexually violent predators. In response to these concerns, the Static-99 developers released revised experience tables based on a large combined sample of 6,406 sexual offenders that now replace the data from the original Static-99 (Hanson & Thornton, 2000). These risk data have been touted as more reliable and representative of sexual offenders in general and sexually violent predators in particular (Harris et al., 2008). This article uses basic psychometric and test construction principles to examine the reliability and validity of new Static-99 experience tables and the proposed interpretation rules as utilized in applied risk assessments with sexually violent predators when determining the legal standard for sexual recidivism. Recommendations for application of the new risk estimates are discussed.

Key words: Static-99, risk assessment, sexually violent predators

Legal schemes to involuntarily confine sexual offenders are premised on the idea that certain sexual offenders are sufficiently dangerous because these individuals have been convicted of one or more qualifying sexual offenses, currently suffer from a legally defined mental disorder or abnormality, and this condition results in these offenders being likely to engage in sexual offenses sometime in the future (risk criterion ; Miller et al., 2005). When the trier of fact determines these conditions are met, individuals being considered for civil commitment are deemed sexually violent predators or sexually dangerous persons (hereinafter referred to as SVP s). The intent of SVP commitment laws is to identify and control what is considered to be a small but extremely dangerous group of sexual offenders upon expiration of their criminal sentence (Janus, 2006; Miller et al., 2005; Woodworth & Kadane, 2004; Doren, 2002).

Laws governing SVP commitment require one or more mental health professionals to conduct an evaluation to determine whether prospective SVP s suffer from a current mental disorder or abnormality and meet the risk criterion (Wollert, 2007a). The risk criterion in many jurisdictions prescribes a certain threshold of risk must be met, usually known as more likely than not or just over a fifty percent probability of sexual reoffending.[1] The probability of meeting the risk criterion is typically assessed by completing an empirically-derived sexual recidivism actuarial risk assessment instrument (SRARA ; Janus & Prentky, 2003; Prentky et al., 2006). SRARA instruments are considered the most accurate method by which to determine the potential for sexual reoffense as compared to unaided clinical judgment, structured professional judgment, or mechanical risk assessment methods (Hanson & Morton Bourgon, 2009; Janus & Prentky, 2003). Moreover, the

methods by which SRARA instruments have been developed allows for transparency in legal proceedings to better understand the true nature of risk assessment, including its significant limits and potential for misuse (Janus & Prentky, 2003). No wonder SRARA has become the sin quo non in predicting the risk criterion in SVP legal proceedings.

Since its dissemination in 1999, the Static-99 has been the most widely used and researched SRARA instrument. It has been; however, criticized on several grounds. The risk percentages (experience tables) from the Static-99 developmental sample (Hanson & Thornton, 2000) have been found unstable when applied to other samples of sexual offenders (Eher et al., 2008; Helmus, 2007a; Saum, 2007; Abracen & Looman, 2006; Looman, 2006; deVogel et al., 2004; Doren, 2004; Langton, 2003). The Static-99 experience tables have also been found to overestimate the probability of sexual recidivism among older age sexual offenders (Barbaree & Blanchard, 2008; Hanson, 2006; Wollert, 2007b). The instability of the Static-99 risk estimates appears to result from differing base rates of sexual recidivism in cross validation studies, where the Static-99 over-predicted sexual recidivism potential in groups with base rates lower than the Static-99 developmental sample and the converse was true in comparison studies having higher base rates (Doren, 2004; Donaldson and Wollert, 2008). These circumstances suggest that the 1,086 sexual offenders comprising the Static-99 developmental sample were not representative of sexual offenders in general or SVP s in particular. New experience tables have been comprised in response to these criticisms. A review of the development of the 2008 experience tables will be presented with a special emphasis on describing risk data in detail and the implication of their application in SVP risk assessments.

Description of the October 2008 Static-99 Experience Tables

Information about the Static-99 2008 experience tables and the application of these data have been released by the Static-99 developers through informal channels of communication such as conference presentations, postings to a web site (www.static99.org), unpublished papers, and, more recently, publication in a professional newsletter (Helmus et al., 2009). To date, none of this information has been subjected to peer review. Despite this fact, clinicians in sexually violent predator cases are relying on information about the new Static-99 experience tables that has been disseminated piecemeal without scrutiny to perform risk assessments and to testify in legal proceedings as to the reliability and validity of the new experience tables. In fact, this situation was a major impetus for writing this paper. The author has provided, to the extent possible, the location where readers may independently obtain unpublished works referenced in this paper. For those documents not readily available by conventional means, interested readers may obtain this information from the author this paper. These references are cited as,   Available upon request from author.

While the October 2008 Static-99 experience tables differ markedly than the sample comprising the original Static-99 risk data, no changes have been made in scoring the Static-99 items (Harris et al., 2003). Each of the ten items on the Static-99 receives a score and the item scores are summed to obtain a total score. The total obtained score corresponds to a score-wise risk level. Score-wise risk levels range between one and ten. The October 2008 Static-99 experience tables do not provide risk data for offenders who obtained total scores of eleven or twelve since none reached these levels. Having four additional score-wise risk levels is a significant change from the original Static-99 risk classification scheme (Hanson & Thornton, 2000) that ranged between zero and six-plus (all scores between six and twelve were collapsed into the six-plus risk category). It is uncertain whether the original qualitative rating system (e.g., low, moderate-low, moderate-high, high) corresponding to the obtained total score from the original data set, as reported by Harris et al., 2003, is applicable in describing the risk potential of members of the new data set (Harris et al.,

2008; Helmus et al., 2009). It is reasonable to assume that the original qualitative rating system is no longer valid for the new experience tables because the range of obtained scores have expanded from seven to eleven levels and the base rate of sexual recidivism has decreased by about one-third in the new data set. Due to these circumstances, the Static-99 developers should recalibrate the qualitative rating system or publish data supporting the validity of cut-scores reported by Harris et al. (2003) continue to be valid with the 2008 experience tables.

At the 2008 annual conference for the Association for Treatment of Sexual Abusers, Harris et al., (2008) released the October 2008 experience tables. The data published to date are available at www.static99.org in a recently released non-peer reviewed article in newsletter (Helmus et al., 2009). The new experience tables contained the samples from Helmus (2008a), as well as adding other studies resulting in a grand total of sixteen samples comprising 6,406 sexual offenders (Complete Sample). The Complete Sample has two derivative groups drawn from it, which are known as the Correctional Services of Canada (CSC) and High Risk members. This leaves a third derivative group, Other Sample, which consists of those offenders who remain after removing the CSC and High Risk groups. The Static-99 developers do not report information about the Other Sample. Major characteristics of the Complete Sample, CSC, and High Risk group are reported in Table 1. The Complete Sample is also subdivided by type of offender (rapists or child molesters).

Table 1: Description of October 2008 Static-99 Samples

Total Sample Size (Complete Sample)	6,406
Number of Samples	17
Average follow up period in years	7.62
Range of follow up period in years	2-16.4
Offender Types- %*	
Child molester	53
Rapist	38
Noncontact	5
Mixed	4
Subsample Sizes- N (% of complete sample)	
High risk	1,273 (19.9)
CSC	1,249 (19.5)
Other	3,218 (60.6)
Rapists	1,747 (27.3)
Child molesters	2,507 (39.1)
Geographic Dispersion-%	
Canadian	32.3
Scandinavian	26.7
Other regions (New Zealand, U.K., & Austria)	21.8

United States	19.2
Recidivism Criterion-%	
Arrest	57.0
Conviction	36.8
Other/Not reported	6.2
Type of Setting-%	
Released from institution	62.0
Community-based	13.8
Mixed	24.2
5-Year Risk Unadjusted Base Rate of Sexual Recidivism-%	
Complete Sample	10.9
High Risk	21.3
CSC	6.6
Other	Not reported
10-Year Risk Unadjusted Base Rate of Sexual Recidivism-%	
Complete Sample	15.8
High Risk	28.1
CSC	11.4
Other	Not Reported

* Based on 10 samples with N = 4,953

Sexual and violent recidivism rates are reported for the Complete Sample, High Risk, and CSC at the five-year and ten-year follow up periods. Violent recidivism data at the five-year and ten-year risk intervals are reported for the Rapist and Child Molester groups because rapists showed higher rates of violent recidivism than child molesters (Harris et al., 2008). The Static-99 developers advise the reported violent recidivism rates for rapists should only be used when assessing for the potential of violent recidivism. Reoffense rates were recorded using life table analysis, fixed follow-up period, and logistic regression. Helmus et al. (2009) recommend using the logistic regression recidivism rates because this statistical method controls for random fluctuations in recidivism rates in groups with small sample sizes. Harris et al. (2008) reported sexual recidivism data on three groups: Complete Sample, CSC, and High Risk. Recidivism data was not reported for the 3,884 sexual offenders comprising the Other Sample.

As seen in Table 1, the samples comprising the October 2008 Static-99 risk data are primarily from Canada and Scandinavian countries (59%) with less than 20% of the members of the sample coming from the United States. Approximately two-thirds of the offenders from the combined studies were released from institutions (prison or hospitals). The exact percentage of offenders who were followed in the community only is at least 14% but may be higher because some studies consisted of a combination of offenders released from prison and those sentenced to disposition in the

community. The proportions of offenders in each category have not been specified. Notably absent from the descriptive statistics are data regarding the age characteristics of the samples. Moreover, data regarding the interaction of age and sexual recidivism have not been reported, although Hanson (2008) stated this analysis is being performed. Unlike the original Static-99 risk data (Hanson & Thornton, 2000) that reported convictions primarily as the measure of sexual recidivism, less than half of the studies (44%) in the current recidivism data were based on convictions only for sexual offenses and one-fourth of the samples used arrests only to classify sexual reoffenses. Four studies (Epperson, 2003; Hanson et al., 2007; Knight & Thornton, 2007; Langton, 2003) recorded sexual recidivism using multiple sources of information. One study did not report the criterion by which sexual recidivism was substantiated (Eher et al., 2008). The observed risk unadjusted base rates in the October 2008 Static-99 experience tables (Harris et al., 2008) are 40% and 30% less as compared to the 2000 developmental sample (Hanson & Thornton, 2000) at the five and ten year follow up periods, respectively.

Logistic Regression and Error Estimates

In summary, the Static-99 developers have advised clinicians to report the logistic regression estimates from the CSC and High Risk groups when presenting results from a Static-99 actuarial risk assessment (Helmus et al., 2009). The Static-99 developers have promulgated a variety of rules or methods to apply and interpret the risk data from the CSC and High Risk groups. This section will review and critique the application of the Static-99 2008 sexual recidivism estimates.

The calculation of the logistic regression recidivism estimates are based on the number of subjects contained in the fixed follow up groups. This results in far fewer subjects comprising the risk groups as compared to the life table analysis method. The different sample sizes for each risk group are contained in Table 2 along with the percentage reduction in the number of subjects followed up using the life table calculation and the fixed follow up period and logistic regression methods. The available number of subjects in the overall Combined Sample and the two derivative samples (CSC and High Risk) contain substantially fewer subjects with the total reduction ranging between 9% and 75%. The reduced number of sexual offenders in the fixed follow up and logistic regression methods of calculating sexual recidivism rates has important implications regarding the accuracy and representativeness of these sample to sexual offenders in general and SVP s in particular. The declining number of subjects in the fixed follow up and logistic regression estimates become increasingly apparent in the small number of subjects at higher risk levels, which is evident in large margins of error in the 95% confidence intervals.

Table 2: Changes in Sample Sizes as Function of Recidivism Calculation

Recidivism Calculation	Complete Sample	High Risk Sample	CSC Sample
Life Table			
5-Year Follow Up	6,406	1,273	1,249
10-Year Follow Up	6,406	1,273	1,249
Fixed & Logistic Regression			
5-Year Follow Up	4,291	1,163	752
10-Year Follow Up	1,621	735	342

% Reduction in Sample Size From Life Table to Fixed & Logistic Regression			
5-Year Follow Up	33.0	8.6	39.8
10-Year Follow Up	74.7	42.6	72.6

The October 2008 logistic regression experience tables (Harris et al., 2008) provide 95% confidence intervals for the risk estimates for each risk score at the five-year and ten-year follow up periods. The inclusion of confidence intervals is an improvement over the original Static-99 risk data (Hanson & Thornton, 2000), which did not report this information. A confidence interval estimates the likelihood that the true value in a group of individuals (a sample drawn from a population, for example, the population of all sexual offenders in the United States) falls within a specific range of values if one were to repeatedly draw independent samples from the same population (Anastasi & Urbina, 1997). The 95% confidence interval reported in the October 2008 Static-99 recidivism estimates (Harris et al., 2008) informs clinicians that the sexual recidivism rates for the sexual offenders comprising each risk level has a 95% probability of falling within the specified range. Alternatively, there is a one in twenty probability that the true recidivism rate for any given risk group reported in the 2008 October Static-99 risk estimates fall above or below the stated range.

Stable risk estimates within each of the October 2008 Static-99 experience tables would be observed when the 95% confidence intervals do not overlap across risk levels (0-10). When confidence intervals from different risk levels do not overlap, we know that the results indicate that one risk level is independent from another. Conversely, when confidence intervals overlap between purportedly independent risk levels, it means there may be no differences in the risk potential between the two groups. Inspection of the October 2008 Static-99 sexual recidivism experience tables reveals that the confidence intervals for all score-wise risk levels within the High Risk and CSC groups overlap and this phenomena occurs to a lesser extent in the Complete Sample. This effect is illustrated in Table 3.

Table 3: Stability of Risk Estimates as Determined by Overlapping 95% Confidence Intervals

Risk Levels With Overlapping Risk Estimates	
5-Year Follow-Up Period	
Complete Sample	0-1; 7-8; 8-9; 9-10
High Risk	0-2; 2-3; 3-4; 4-5; 5-6; 6-7; 7-8; 8-10
CSC	1-3; 2-4; 3-5; 4-6; 5-7; 6-9; 7-10
10-Year Follow-Up Period	
Complete Sample	0-1; 1-2; 2-3; 7-8; 8-9; 9-10
High Risk	0-2; 3-4; 4-5; 6-8; 7-9; 8-10
CSC	0-3; 2-4; 3-5; 5-7; 6-9; 7-10

Inspection of the data in Table 3 indicates that the Complete Sample has the most stable risk estimates at the five year follow up period as compared to the CSC and High Risk subsamples. The High Risk and CSC subsamples show significant instability in risk estimates across all score-wise risk levels, with the greatest amount observed within the CSC group. In some instances, confidence intervals overlap between three or four score-wise risk levels. This trend appears to result from decreasing sample sizes at each risk level. Consequently, the lack of independence between score groups does not support having eleven risk levels (0-10) but rather it appears the Static-99 developers should conduct further analysis to reduce the number of score groups in an attempt to achieve statistical independence or, at least, reduce the degree of overlap between score groups.

Examination of the 95% confidence intervals across samples determines to what extent the CSC and High Risk samples are independent of one another at the five-year and ten-year follow up periods. A comparison of the CSC and High risk groups with the Complete Sample was not possible because this group is not independent of the CSC and High Risk groups as each derivative sample is also contained in the Complete Sample numbers. Assuming that the CSC and High Risk groups are considered independent groups, an analysis of the overlap of confidence intervals was conducted based on a formula proposed by Cumming and Finch (2003). With few exceptions, this analysis showed that the CSC and High Risk groups appear independent of one another in terms of risk potential through score-wise risk level eight. The analysis for the remaining two high risk scores, nine and ten, could not be conducted because the total number of subjects were too few to meet the assumptions to apply the formula testing for independence of samples (Cumming & Finch, 2003).

Application of Logistic Regression and Error Estimates in SVP Proceedings

Despite the apparent independence between the CSC and High Risk groups, the combination of the overall low base rates of sexual recidivism and the small sample sizes at higher score-wise risk levels (8-10) raises a practical dilemma for clinicians opining about the risk criterion in states where the SVP risk criterion is based on a standard of slightly over fifty percent (i.e., more likely than not). Clinicians assessing SVP s will find it difficult to affirm the more likely than not risk criterion when utilizing the CSC and High Risk logistic regression experience tables as recommended by the Static-99 developers (Harris et al., 2008). This circumstance is explained below.

The logistic regression estimates found in the studies underlying the CSC sample never reach fifty percent at any score-wise risk level at either the five-year or ten-year follow up periods. On the other hand, the more likely than not standard could be substantiated by clinicians when relying on the upper limit of the 95% confidence interval of the CSC sample at score-wise risk level ten (five-year follow up) and score-wise risk levels nine and ten (ten-year follow up). In each of these circumstances; however, the lower limit of each 95% confidence interval falls well below 50%. Inspection of the logistic regression data for the studies comprising the High Risk sample indicate the more likely than not standard is met by the logistic regression estimates at score-wise risk levels nine and ten at the ten-year interval and at score-wise risk level ten at the five year follow up period. In each of these instances; however, the lower bound of the 95% confidence interval falls below 50%. The upper bound of the 95% confidence interval for the score-wise risk level nine at the five-year follow-up and the score-wise risk level eight at the ten-year interval exceed 50% but the logistic regression estimates and lower limits of the 95% confidence intervals fall below more likely than not.

The instability of the logistic regression risk estimates in the CSC and High Risk groups in predicting the SVP risk criterion is similar when compared to the Complete Sample. The logistic regression

estimates and the upper bounds of the 95% confidence intervals for the studies constituting the Complete Sample do not meet the more likely than not standard for score-wise risk levels between zero and nine at the five-year follow up and between zero and seven at the ten-year risk interval. When considering the upper bounds of the 95% confidence intervals, the more likely than not criteria is met at met at score-wise level ten at the five-year follow up period but the lower limit of the 95% confidence interval falls below 50%. At the ten year risk interval, the lower limit of the 95% confidence interval for the score-wise risk level eight falls below chance levels. At the score-wise risk levels of nine and ten for the ten-year follow-up, the lower limit of the 95% confidence intervals do not fall below 50%, which suggests these score-wise risk levels are reliable in substantiating the more likely than not risk criterion. Despite this fact, the Static-99 developers advise clinicians to use only the CSC and High Risk samples as reference groups in which compare individuals being assessed.

When considering the lower limit of the margins of error associated with the logistic regression estimates in the CSC and High Risk groups, no score-wise risk levels at either five or ten years meets the more likely than not standard. The degree of measurement error in these two groups makes it difficult for clinicians to reject the null hypothesis (i.e., subject being assessed does not meet the risk criterion). Wollert (2006) explains this problem:

> For an expert to be reasonably certain in rejecting the null hypothesis that the recidivism risk for a respondent is not the same as the risk for non-SVPs, he or she must be reasonably certain that the lowest plausible estimate of the respondent s risk level exceeds the non-SVP standard. The lowest plausible estimate for a respondent must always be less than the respondent s obtained test score, however. This difference, along with the width of the corresponding confidence interval (or what might also be called the region of doubt), is due to measurement error, which arises because experts sometimes disagree when they score the same group of subjects on the same test. If the measurement error for a test is small, the region of possible detection error will be narrow, and the lowest plausible estimate will not fall too far below the level suggested by the offender s obtained test score. If it is large, the region of possible error will be wide, and the lowest plausible estimate will be substantially less than the level suggested by the offender s obtained test score. It is therefore important for experts to consider measurement error when deriving predictions because the chances of rejecting the null hypothesis decrease as measurement error increases (p. 79).

It is of concern that the Static-99 test developers have provided interpretation rules that appear to avoid or work around the issue of the CSC and High Risk logistic regression estimates failing to substantiate the more likely than not criterion. The developers of the October 2008 Static-99 experience tables devised a report template for clinicians to document the risk data from the High Risk and CSC samples in reports or testimony, which has recently been replaced with revised reporting procedures (Helmus et al., 2009).[2] The original reporting methods instructed clinicians to report the lower bound of the 95% confidence interval for the CSC group and the upper limit of the corresponding 95% confidence interval for the High Risk sample consistent with the individual s score-wise risk level. Failing to specify the opposing bound of the 95% confidence interval is considered contrary to standard practice in reporting confidence intervals (AREA, 1999; Cumming & Finch, 2005). Inspection of the 2008 experience tables for the CSC group shows that the more likely than not criterion in SVP statutes cannot be met when reporting the lower limit of the 95% confidence interval at any score-wise risk level. The more likely than not standard can be met when

reporting the upper limit of the 95% confidence interval from the High Risk sample at score-wise risk levels eight, nine, and ten at the ten-year follow-up period. Yet, the more likely than not threshold would not be substantiated if clinicians considered the lower limit of the 95% confidence interval at these same scores. It becomes apparent that this method of reporting confidence intervals obscured the fact that the more likely than not standard cannot be met when considering the risk data from the CSC and High Risk samples. This same obfuscation occurs in the revised reporting procedures.

The new reporting guidelines (Helmus et al., 2009) abandon the use of 95% confidence intervals altogether. They instruct clinicians to report only the logistic regression estimates from the CSC and High Risk groups at the score-wise risk level assigned to the individual being assessed. This method is contrary to accepted standards of reporting data with associated error measurement (AREA, 1999; Cumming & Finch, 2005), in this case the 95% confidence interval. Failing to describe the potential error in the sexual recidivism rates robs the trier of fact from knowing that the lower limit of the 95% confidence interval falls below the more likely than not criterion for those logistic regression estimates at or above fifty percent within the CSC or High Risk groups. Consequently, the trier of fact is left with the impression that the individual being assessed meets the more likely than not criterion when the clinician cannot be reasonably certain of this fact if the lower limit of the 95% confidence interval was considered and reported (Wollert, 2006). The current reporting procedures also introduce clinical judgment into the actuarial risk assessment method by having clinicians determine which logistic regression estimates, CSC or High Risk, best approximate the risk potential of the individual being assessed. Helmus et al. (2009) concede this instruction requires clinicians to combine actuarial science with clinical judgment and acknowledge there is no empirical research available to assess how well evaluators can make this judgment. Introducing this potential source of error into the risk assessment result, which is unquantifiable and unknown, is unacceptable in a forensic context where the trier of fact needs to know the reliability of the risk estimates to assign appropriate weight to the evidence in deciding whether the individual meets the SVP risk criterion. Psychologists who follow the Static-99 reporting procedures (Helmus et al., 2009) without properly qualifying opinions are likely to run afoul of ethical principles for psychologists (American Psychological Association, 2002) regarding the obligation to report limitations in the bases of assessments.

Generalizability of the Logistic Regression Estimates

In the previous section, it was shown that the logistic regression estimates for the CSC and High Risk samples are of dubious reliability when assessing the more likely than not standard in SVP proceedings. This is not the sole basis by which to determine the applicability of the October 2008 Static-99 logistic regression estimates to other populations of sex offender in general or SVP s in particular. Clinicians must also consider the relevancy or fit of the new Static-99 risk data to other populations of offenders, which is also known as generalizability theory (Brennan, 2001). The extent to which the observations of a study sample (Static-99 CSC and High Risk groups) can be considered applicable to a larger population (e.g., all sexual offenders) or a specific universe (all SVP s) hinges upon the extent to which the study sample is representative of the comparison groups. The October 2008 Static-99 experience tables have been touted as being more representative of sexual offenders in general than the predecessor risk data from the 2000 Static-99 developmental sample (Hanson & Thornton, 2000) based on the greater number of sexual offenders comprising the new data set (Harris et al., 2008). This validity of this assertion will be examined.

The developers of the October 2008 Static-99 experience tables refer to the data as norms. Norms is a term of art in statistics that refers to data from a sample of individuals that can be used to

describe the typical performance of members of a larger population from which the sample was drawn (Anastasi & Urbina, 1997). In other words, norms can be relied upon as being representative of the larger population from which the normative group was sampled. Adult intelligence test scores based on the Wechsler Adult Intelligence Scales: Third Edition (WAIS-III ; Tulsky et al., 2002) are a commonly recognized form of normative data. The WAIS-III assesses intellectual functioning of adults between the ages of sixteen and eighty-nine in the United States (Tulsky et al., 2002). The WAIS-III standardization sample was obtained using a weighted sampling method that ensured the group of 2,450 adults matched 1995 United States census data as closely as possible based on the variables of age, gender, ethnicity, geographic location, and educational levels. The WAIS-III was administered to the standardization sample and the results were used to develop normative data for various forms and levels of intellectual functioning for adults between the ages of sixteen and eighty-nine in the United States. Sampling methods, like those used in the development of the WAIS-III, ensure the results from the sample best approximate the performance of members from the larger population (Kazdin, 2003; Anastasi & Urbina, 1997; & Kalton, 1983).

The methods by which samples were drawn to constitute the October 2008 Static-99 experience tables stand in stark contrast to random sampling techniques used in the development of normative data. The Static-99 developers did not attempt to define a population of sexual offenders by certain characteristics and then sample from this population to obtain a representative standardization sample. Other random sampling techniques were not used to reduce the potential of error influencing the results of the obtain risk estimates. For instance, none of the underlying sixteen studies comprising the experience tables utilized random sampling techniques to obtain members of the samples. The Static-99 developers did not randomly assign subjects from the sixteen studies to the Complete Sample or its derivative groups (CSC & High Risk). The 6,406 members of the Complete Sample and the subjects comprising the two derivative groups (High Risk and CSC) were chosen based on convenience rather than being representative of sexual offenders in general or SVP s from the United States in particular. In essence, the developers of the risk data for the October 2008 Static-99 experience tables appear to assert that the international group of sexual offenders making up the sixteen samples represent normative behavior of a larger population of sexual offenders without showing how these samples match characteristics of a larger universe of sexual offenders in general or SVP s in particular. This circumstance is a serious deficiency in the October 2008 Static-99 experience tables, which has not been corrected from the original Static-99 risk data (Hanson & Thornton, 2000).

The developers of the new Static-99 risk data did not seek to define characteristics of an international population of sexual offenders or SVP s from the United States. Without defining the characteristics of a specific sex offender population and not taking into account their heterogeneity on such variables as age, ethnicity, geographic location, income levels, types of sexual offenders (e.g. child molesters, rapists, noncontact, or mixed), and type and length of sentence, the developers of the October 2008 Static-99 risk data cannot ensure the experience tables adequately represent the risk potential of sexual offenders in general. Without defining additional parameters related to the number of prior sexual crimes and presence and type of mental disorder or abnormality, the application of the October 2008 risk data to SVP s remains elusive. Consequently, clinicians cannot assume the October 2008 Static-99 experience tables represent normative sexual recidivism behavior of sexual offenders in general or SVP s in particular.

A corollary to the issue about reference group representativeness is illustrated by examining the Other Group, those who did not fit the High Risk or CSC samples. The Static-99 developers have published little information about the 3,884 offenders comprising the Other Group. Four of the seven studies comprising the Other Sample (Eher et al., 2008; Allan et al., 2007; Langstrom, 2004; Bartosh et al., 2003; Epperson, 2003) contain at least 2,663 prison releasees while one study

(Harkins & Beech, 2004) consisted of offenders who were released from prison and sentenced to dispositions in the community with no ability to discern the former from the latter. The Other Sample contains the largest number of prison releasees as compared to the High Risk or CSC groups. This circumstance is important because individuals eligible for release from prison are subject to provisions of civil commitment schemes. Thus, this group is likely more comparable to SVP s than the High Risk or CSC samples. The Static-99 developers should publish more information about the offender, treatment, and methodological characteristics of this group to determine if this would be the most appropriate reference group in which to compare individuals being assessed under SVP statutes.

Current Proposed Interpretation Rules

The developers of the October 2008 Static-99 experience tables devised methods by which to document risk data (Helmus et al., 2009). The published instructions are supplemented with additional information contained on the Static99.org website. These reference materials advise clinicians how to report the risk data from the CSC and High Risk experience tables when describing the risk potential of a particular individual, how to select the most appropriate reference group for the person being assessed, and the method by which to report relative ranking of risk from a third group referred herein as the Percentile Rank group. The following section will review and discuss the efficacy of the currently available interpretation rules.

What Are the Current Interpretation Rules?

Doren and Thornton (2008) and Helmus et al. (2009) have proposed certain guidelines for determining whether to compare an individual to the High Risk or CSC samples. They propose using the High Risk group when an individual is resistant to sustained rehabilitative efforts, has been expelled from treatment, has dropped out from treatment, self-reports sexual deviancy, demonstrates increased salient dynamic factors, and/or exhibits recent antisocial behavior during their current sentence. Doren and Thornton (2008) recommend utilizing the CSC risk information under the conditions of the individual participating in limited programming, receiving treatment consistent with the risk, need, and responsivity model (Andrews & Bonta, 2006), and/or has been cooperative but proportionate programming is not provided. To this list of possible discriminating factors, Helmus et al. (2009) add that the sexual offenders were gradually reintegrated into the community by parole and human services programming. It is entirely unclear whether clinicians must substantiate all or some of the proposed selection conditions when determining the most comparable reference group for a particular offender or how to reconcile situations where an offender may meet conditions consistent with both groups. Doren and Thornton (2008) and Helmus et al. (2009) provide no empirical support as to whether the proposed selection conditions differentiate individuals from the two risk groups at a statistically significant level, how these criteria were developed, or how these criteria are objectively measured in a standardized way.

The current Static-99 report template (available at the Static99.org website) suggests that individuals who have been determined to meet the criteria for civil commitment are most similar to the High Risk sample based on the fact that such individuals have more risk factors external to the Static-99 than the typical sexual offender. This statement is unclear in two ways. First, it does not specify whether a civil commitment determination must be made by a trier of fact or as a result of a diagnostic opinion rendered by a state appointed evaluator. If the former situation is true, the Static-99 developers provide no guidance on what risk group an evaluator should use for individuals undergoing initial civil commitment proceedings. Second, the Static-99 developers assert certain, unspecified risk factors outside the realm of the Static-99 are endemic to and distinguish SVP s

from other types of sexual offenders. Nowhere in the published information do they identify these risk factors or show the statistical properties supporting the discriminating power of these variables. Consequently, it cannot be reliably determined that these unnamed variable are characteristic of the High Risk group as compared to the CSC, Other Sample, and Percentile Rank group. As will be discussed later, using the High Risk sample as the reference group for SVP s will result in unacceptably high erroneous decisions that individuals meet the risk criterion.

Phenix and Arnold (2008) endorse the selection methods promulgated by the Static-99 report template but, also, propose that clinicians consider other factors to establish where the individual falls within a specified range of risk, where the floor is the score-wise logistic regression average for the CSC group and the ceiling is the score-wise logistic regression average for the High Risk group. Phenix and Arnold (2008) contend that weighing these factors can determine where an individual falls within the lower, middle, or upper levels of the logistic average range. They advise clinicians to consider four variables (Psychopathy Checklist Revised (PCL-R) score, treatment drop-out or completion, ongoing antisocial behavior, and lack of compliance with supervision in the last two years) in combination with the score-wise risk levels from the Static-2002, MnSOST-R, and/or SORAG. By weighing up to seven factors along with the Static-99 results in some unspecified manner, Phenix and Arnold (2008) reason this approach will improve the accuracy of the risk prediction for an individual. There are five problems with the Phenix and Arnold (2008) method that make it unreliable and unworkable in its current form.

First, Phenix and Arnold (2008) provide no instructions for clinicians to quantify and measure treatment drop-out or completion, ongoing antisocial behavior, and lack of compliance with supervision. Cut-off scores for the PCL-R are not prescribed to correspond to the risk rankings within the specified bounds of the logistic regression averages. Lack of standardized scoring instructions reduces the reliability in measuring the four variables and increases the potential for making erroneous rankings within the score-wise range of logistic regression averages (Anastasi & Urbina, 1997). Second, Phenix and Arnold (2008) fail to report data on the PCL-R scores, treatment drop-out and completion rates, ongoing antisocial behavior, and lack of compliance with supervision for the CSC and High Risk groups. This data would be necessary to develop empirically based selection criteria that would allow clinicians to determine what mix of scores on the four variables would correspond to the appropriate ranking (lower, middle, upper) within the range of score-wise logistic regression averages. Third, it is uncertain whether an individual has to meet all or some of the four conditions specified to be assigned the rank of low, middle, or high. There may be situations where prospective SVP s have more than one but less than four of the factors in which case a clinician has no guidance as to the appropriate rank to choose. Phenix and Arnold (2008) provide no solution to this dilemma. Fourth, the inclusion of other actuarial risk assessment findings is of dubious validity. It is not enough to advise clinicians to consider score-wise risk levels from the MnSOST-R, Static-2002, or SORAG without providing empirically supported methods to do so. Data has not been reported for the score-wise risk levels for members of the Static-99 CSC or High Risk groups on the MnSOST-R, Static-2002, or SORAG. It is impossible to devise a valid system by which to select the appropriate risk-rank within the range of score-wise logistic regression averages when no data exits as to what combination of cut-scores from three or four actuarial instruments reliably discriminates the three risk levels. Finally, Phenix and Arnold (2008) fail to propose a reliable solution for clinicians to combine the four factors and various combinations of score-wise risk levels from multiple actuarial instruments to pick the most valid risk rank within the score-wise range of logistic regression averages.

Clinicians should be wary of using the selection criteria proposed by Doren and Thornton (2008), and Helmus et al. (2009) for determining the representativeness of the Static-99 reference groups to individual sexual offenders that have not been adequately described in terms of the characteristics

of the samples and documented in a standardized format, and whose discriminative power remains unproven. Vague and ambiguous terms to determine the applicable reference group to compare an individual sexual offender (Doren & Thornton, 2008; Helmus et al., 2009) or to rank an offender s level of risk with a specified range (Phenix & Arnold, 2008) leave the door open to these conditions being interpreted and applied in as many ways as the number of clinicians who use them. In absence of specific formulas to make these determinations, this approach is pseudo-actuarial enabling confirmatory bias to parade around under the guise of actuarial judgment.

Moreover, the recommended interpretative approaches propounded by Doren and Thornton (2008), Helmus et al., 2009, and Phenix and Arnold (2008) increase the measurement error of the recidivism rates to an unspecified degree. This makes it impossible to determine the reliability of the scoring system.[3] It is incumbent upon the Static-99 developers to provide sufficient data about decision rules (AREA, 2003) that results in the necessary transparency for clinicians to defend the selection of reference groups on which they rely to estimate the potential of an SVP meeting the risk criterion (Janus & Prentky, 2003). Instead, clinicians must rely on blind faith in the Static-99 developers unsubstantiated assertions that the selection criteria are reliable and valid in deciding which reference group is most comparable to an SVP who is being assessed under the risk criterion. This is an unsettling proposition when exposing the currently proposed interpretation rules to the light of empirical research.

The Discriminating Power of Treatment

Doren (2008) and Doren and Thornton (2008) assert that contemporary treatment methods have not only caused the base rate of sexual recidivism to drop from the Static-99 2000 developmental group risk data (Hanson & Thornton, 2000) as compared to the October 2008 Static-99 experience table for the Complete Sample but they also contend that participation in treatment has sufficient discriminative power to allow clinicians to determine whether an individual should be compared to the High Risk or CSC samples, a proposition also endorsed by Phenix and Arnold (2008). This interpretative guideline appears speculative at best since neither Doren (2008) or Doren and Thornton (2008) provide empirical data supporting it. Determining the treatment status of members of the October 2008 Static-99 experience tables is not only confusing but empirical results (Losel & Schmucker, 2005; Marques et al., 2005; Rice & Harris, 2003; & Furby et al., 1989) do not support the discriminating power of treatment participation as a rationale to select a reference group in which to compare the risk potential of a prospective SVP under the risk criterion.

The data describing treatment of the sexual offenders in the October 2008 Static-99 offender group is described in a chart entitled Static-99 Replications: Descriptive Information (Harris et al., 2008). The treatment status of the subjects from the combined studies is classified under the heading Mostly Treated? The following provides a summary of the treatment status classification with the percentage of each category in parenthesis: Yes (20.0%), Mixed (33.4%), and Not Reported (46.6%). No information is provided as to the Not Reported category. It is uncertain whether this category reflects those offenders who were not treated, or if the treatment status of offenders in these studies are uncertain, or a combination of both. The method by which the Static-99 developers (Harris et al., 2008) summarized the treatment status of the offenders suggests that reliable data does not exist for each member of Complete Sample. Within the studies where treatment was administered, no data is reported as to the proportion of offenders who completed treatment or dropped out. No information is provided as to what proportion of the untreated offenders were individuals who refused treatment. These circumstances are important to know because it has been documented that treatment dropouts are at higher risk than treatment completers to reoffend sexually (Hanson et al., 2002; Rice & Harris, 2003, & Losel & Schmucker, 2005) and treatment refusers show a lower likelihood to reoffend sexually than treatment

completers (Losel & Schmucker, 2005). Lacking analysis about these important treatment variables for the offenders constituting the October 2008 Static-99 experience tables calls into question the efficacy of using treatment as an interpretation rule for choosing which sample (High Risk vs. CSC) to compare an individual being assessed.

In defense of using treatment as a selection criteria, Doren (2008) reasons that the effect of treatment for those sexual offenders within the Complete Sample in the October 2008 Static-99 risk data account for the 6.5 percentage point reduction in the risk unadjusted base rate of sexual recidivism at the five-year follow up period as compared in the original Static-99 developmental sample (Hanson & Thornton, 2000; 17.4% - 10.9% = 6.5%). To further buttress his argument, Doren (2008) cites a meta-analysis of sex offender treatment outcome studies conducted by Losel and Schmucker (2005) where the treated sexual offenders had a sexual recidivism rate that was 6.4 percentage points lower than the untreated offenders. Doren (2008) concludes the similarity in the absolute decreases in sexual recidivism rates between the two studies is proof positive that the effects of treatment among the members of the October 2008 Static-99 Complete Sample lowered the base rate of sexual recidivism from the data reported in the original Static-99 (Hanson & Thornton, 2000). Other than to base his conclusion on circular reasoning, Doren (2008) fails to produce empirical evidence that the recidivism rates between samples studied by Hanson and Thornton (2000) and Harris et al., (2008) were actually statistically significant or caused by treatment effects. This is an important determination to verify because the 6.4 percentage point difference between the treated and untreated sexual offenders found by Losel and Schmucker (2005) was considered to be statistically nonsignificant because moderator variables appeared to interact with the effect of treatment in reducing sexual recidivism rates.

Losel & Schmucker (2005) conducted additional statistical analysis in the studies reviewed to uncover what treatment, offender, and methodological characteristics contributed to the treatment effect size difference showing that treated sexual offenders as a group had a lower aggregate recidivism rate than the control group of untreated sexual offenders. This analysis was conducted using Odds Ratios. The following summarizes results most relevant to the October 2008 Static-99 experience tables. Most notably was the fact that sexual offenders who underwent bilateral orchiectomy not only had recidivism rates lower than offenders who received other forms of treatment (hormonal, cognitive, or behavioral therapies), but physical castration contributed a large proportion of the variance in the effect size for the treated group. Effect sizes were significantly greater for sexual offenders who volunteered for treatment as compared to those subjected to involuntary treatment. In situations when the authors of studies were also involved in dispensing treatment, significantly larger effect sizes were observed. Treatment drop outs reoffended at higher rates than treatment completers. Yet, on the other hand, untreated comparison groups comprised of treatment refusers actually had slightly lower recidivism rates than treatment completers. Finally, offenders who participated in outpatient treatment programs had lower sexual recidivism rates than those offenders who received therapy in institutions (prisons and hospitals). Harris et al. (2008) produced no data about these variables.

As the analysis by Losel and Schmucker (2008) and others (Marques et al., 2005; Hanson et al., 2002; Harris & Rice, 2003; & Furby et al., 1998) reveal, the simple dichotomy of treated and untreated sexual offenders appears insufficient to discriminate the recidivism behavior of these groups. In light of this fact and absent empirical analysis, it is premature for the developers of the October 2008 Static-99 experience tables to advise using treatment participation as a major determinate by which to select a Static-99 reference group to compare an individual undergoing an actuarial risk assessment. The developers of the October 2008 Static-99 experience tables should at least conduct analyses of the Complete Sample and its derivative groups on the various treatment, offender, and methodological variables Losel and Schmucker (2005) found to influence

recidivism rates between the treated and untreated sexual offenders to determine what, if any, treatment variables should be considered in deciding what Static-99 reference group is best representative of an individual being assessed.

Antisocial Behavior as a Distinguishing Characteristic

Helmus et al. (2009) and Doren and Thornton (2008) state that clinicians should consider the High Risk reference group to compare the recidivism potential of sexual offenders who exhibited antisocial behavior during their most recent sentence. No data has been presented to authenticate how this factor was developed and its statistical properties in discriminating High Risk offenders from the CSC or Complete Sample groups. This creates two major problems for clinicians who attempt to justify what reference group they choose to compare an individual offender. First, the Static-99 developers offer no statistical analysis as to   antisocial behavior during current sentence being a reliable factor by which to distinguish sexual offenders in the CSC sample from the High Risk group. Second, the Static-99 developers provide no guidance as to how to quantify or measure antisocial behavior during current sentence   in a reliable or valid fashion. This will place clinicians in a situation where they will rely on idiosyncratic criteria to confirm or disconfirm the presence of this factor in selecting a Static-99 reference group. These circumstances clearly conflict with the idea of a SRARA instrument providing sufficient transparency to allow triers of act to assign appropriate weight to risk assessment opinions (Janus & Prentky, 2003). Rather, using ambiguous, nonstandardized terms with pejorative connotations and lacking an empirical basis regarding the discriminative power of this interpretation rule will result in conflicting expert opinion testimony that will only serve to confuse the trier of fact.

Similar to using the construct of antisocial behavior to discriminate reference group selection, Phenix and Arnold (2008) propose using a PCL-R cut-off score of ≥ 30 to increase an individual s risk level to the high range on the Static-99 presumably even when an obtained raw score on the test indicates a lower risk level. The reasoning behind this interpretation rule is unsound for two reasons. First, it is recognized that PCL-R scores are weakly and inconsistently related to predicting future sexually reoffending behavior (Hare, 2003). Phenix and Arnold (2008) provide not data to contradict this finding. Second, using other risk factors to adjust the actuarial risk prediction as proposed by Phenix and Arnold (2008) amounts to combining clinical judgment with actuarial science, thus decreasing the reliability of the risk assessment (Janus & Prentky, 2003), as well as lowering predictive accuracy of sexual recidivism estimates (Hanson & Morton-Bourgon, 2009).

Reporting Percentile Ranks

Helmus et al. (2009) advise clinicians to report percentile ranks and the Static-99 developers supply a percentile rank table[4] and reporting procedures for this purpose.[5] Percentiles ranks transform a obtained total score distribution into a distribution of standardized scores that represent the proportion of scores in a distribution that a specific score is greater than, less than, or equal to (Stockburger, 1998). The transformation of obtained total scores into percentile ranks destroys the ability to determine the proportion of difference between scores (Stockburger, 1998). Consequently, a clinician relying on percentile ranks cannot state that a specific percentile rank score meets, exceeds, or falls below the risk criterion established by law.

The instructions for using percentile ranks are especially prejudicial toward individuals undergoing SVP evaluations at the high score-wise risk levels. Individuals who are subjected to civil commitment trials typically have high obtained total scores on the Static-99. As discussed previously, the more likely than not standard cannot be met at the five and ten year follow-up periods for the CSC and High Risk groups. At score-wise risk levels between five and ten at the

ten-year follow-up, the corresponding percentile ranks range between 74.9 and 100 yet the percentage of risk for sexual reoffense falls below 50% either at the average logistic regression estimate or at the lower bounds of the 95% confidence interval (assuming the percentile ranks table represents values at the ten-year risk interval). Consequently, triers of fact will be confronted with what appears to be conflicting risk information and may reconcile this difference in favor of determining the percentile ranks as being a more accurate reflection of risk than the absolute level of risk inferred by the logistic regression estimates. The misattribution of high risk implied by reporting percentile ranks may be sufficient to sway triers of fact to find an individual meets the SVP criteria due   to great pressure to lock up sexual offenders and the difficulty of the trier-of-fact [judge or jury] has in understanding the basis of   sophisticated   professional judgment   (Campbell, 2007).

A final note regarding the use of percentile ranks is in order. A percentile indicates the individual s position in a normative or standardization sample (Anastasi & Urbina, 1997), in this case the relative ranking of a person undergoing SVP proceedings as compared to a group of Canadian sexual offenders only. It is unknown to what extent the percentile rank group is comprised of members from the CSC and High Risk samples from which the logistic regression estimates are reported. As discussed earlier in this report, it cannot be assumed that the Static-99 2008 experience tables accurately portray the normative risk potential of sexual offenders in general from the United States or SVP s in particular. This problem becomes magnified when the comparison group is restricted further to Canadian sexual offenders only. It raises an important question about whether a group of Canadian sexual offenders are representative of the relative risk potential of SVP s in the United States. Since individuals who come under SVP laws are considered high risk sexual offenders, it is reasonable to conclude this preselected population will score higher on the Static-99 than a general sexual offender population from Canada. Persons undergoing SVP evaluations will be prejudiced by this bias in that SVP s will consistently have higher and restricted range of score-wise risk levels (e.g., moderate to high) as compared to members of the Canadian percentile rank group (e.g., low to high). The restriction in score variance found in the SVP group makes for an unequal comparison to the Canadian percentile rank group and creates an artificial inflation of SVP s percentile standing. In the final analysis, the Static-99 developers   recommended method for reporting of percentile ranks place clinicians in the position of misinforming triers of fact that the reported percentile ranks reflect the relative standing of the individual as compared to the CSC and High Risk groups and presumes, without empirical substantiation, the risk potential of the Canadian only group is representative of individuals undergoing SVP trials. The high likelihood of confusion when trying to untangle the reasons for reporting logistic regression data for two groups and the percentile ranks from a third Canadian only sample far outweighs the probative value to a trier of fact or, for that matter, a clinician in determining whether an individual meets the SVP risk criterion. In fact, the reporting of percentile rank data as prescribed by the Static-99 test developers is wholly irrelevant in determining the risk standard under the law.

In summary, the reporting of the 2008 Static-99 percentile ranks is misleading because this data has no relevancy or relationship in determining whether a prospective SVP meets the risk criterion. The prejudicial impact of percentile ranks outweighs any probative value in assisting the trier of fact to decide the risk criterion. In fact, the percentile ranks provided by the Static-99 test developers are discriminatory when applied to prospective SVP s, as their relative ranking of risk will be consistently higher in comparison to the Percentile Rank group comprised of lower risk sex offenders. Psychologists are obligated not to use scores from instruments that are considered discriminatory (AREA, 2003). Considering these circumstances, clinicians should not report or testify about percentile ranks when describing the risk data for SVP s derived from the October 2008 Static-99 experience tables.

Base Rates Matter!

The Static-99 developers (Helmus et al., 2009) declared that base rates of sexual recidivism matter as seen in the significant decline in sexual recidivism between the 2000 risk data (Hanson & Thornton, 2000) and the October 2008 experience tables. Despite this fact, the base rates of sexual recidivism published by various studies comprising the Static-99 experience tables for the Complete Sample, High Risk, and CSC (see Table 1) are likely to be inaccurate when applied to offenders who are older than forty and in locales where established base rates are lower or higher than what was found in the October 2008 Static-99 risk data pool. Clinicians should be aware of and consider variations in local sexual recidivism base rates where available (AERA, 2003) when rendering opinions as to whether a prospective SVP meets the statutorily defined risk criterion. Two examples illustrate this consideration.

The October 2008 Static experience tables do not account for the mitigating affect of advancing age on the reduction of sexual recidivism. Many studies have found that sexual recidivism rates decline with advancing age when accounting for offender types (Hanson, 2002 & Prentky & Lee, 2007), when combining offender types into one group (Barbaree & Blanchard, 2008; Fazel et al., 2007; Knight & Thornton, 2007; Milloy, 2007; Hanson, 2006; Thornton, 2006; Hanson & Morton Bourgon, 2004; Harris & Hanson, 2004; Langstrom et al., 2004; Sjostedt & Grann, 2002; Sjostedt & Langstrom, 2001; Hanson & Bussiere, 1998; & Song & Lieb, 1995), or after controlling for age at release confounds in actuarial risk items and total actuarial scores (Barbaree et al., 2007; Barbaree et al, 2009). These data generally suggest that rates of sexual recidivism reduce as a function of advancing age, with the steepest declines occurring in the age decades of fifty, sixty, and beyond. Likely explanations for these trends are beyond the scope of this paper but interested readers are referred to Barbaree and Blanchard (2008). Hanson (2008) indicated that the Static-99 developers plan to complete an analysis of the affect of age on sexual recidivism rates in the October 2008 experience tables. Until the Static-99 developers complete this analysis, clinicians should exercise caution when utilizing the October 2008 Static-99 experience tables with sexual offenders who are age forty or older. In the interim, clinicians may want to use the age adjusted risk information released by Hanson (2006) as corrected by Waggonner et al., (2008) or employ the age correction methods proposed by Wollert (2006 & 2007b) when considering the affect of advancing age in reducing the risk of sexual recidivism for SVP s.

The base rate of sexual recidivism for the High Risk group is 21.9% at the five-year follow up and 29.8% at the ten-year interval, which is greater than the original Static-99 base rates (Hanson & Thornton, 2000) for comparable follow-up periods (17.4% at five years and 22.7% at ten years). This is not surprising as the High Risk sample consists primarily of studies where the sexual recidivism was sampled decades in the past and, in one instance, up to forty or fifty years ago (Knight & Thornton, 2007). Moreover, almost three-fourths of the offenders were considered to be highly disturbed hospitalized offenders who were known to have high rates of sexual recidivism. The base rates of sexual recidivism in the High Risk group are contrary to much lower base rates of sexual recidivism found in more contemporary samples of sexual offenders from the United States that range between 3% and 12% for sexual offenders released from prison or sentenced to community dispositions (California Sexual Offender Management Board, 2008a, b, c; Minnesota Department of Corrections, 2007; State of Delaware, 2007; State of New York, 2007; Tennessee Bureau Investigations, 2007; Institute of Public Policy, 2006; Washington State Institute of Public Policy, 2005; Wyoming Legislative Service Office, 2005; Langan et al., 2003; State of Ohio, 2001; Iowa Department of Human Rights, 2000; Arizona Department of Corrections, 1998; & Song & Lieb, 1995) and 4.3% sexually violent predatory recidivism rate observed in a group of certified SVP s who were released from commitment without having completed or participated in treatment (Padilla, 2006). As demonstrated by Donaldson and Wollert (2008), sexual recidivism percentages in

actuarial instruments that are predicated on base rates higher than a local population of sexual offenders will overpredict the rate of sexual recidivism in the local group and results in high rates of false positive opinions that respondents meet the SVP risk criterion.

Consequently, clinicians should compare local base rates, where available, with the risk unadjusted base rate of the selected Static-99 reference group. To the extent that the local base rate is less or greater than what was found in the Static-99 reference group, clinicians should qualify the risk opinion by stating that the Static-99 reference group likely overpredicts (local base rate is lower than the Static-99 reference group) or underpredicts (local base rate is higher than that found in the Static-99 reference group) the recidivism potential of the SVP being assessed. Alternatively, clinicians can employ the base rate correction procedures presented by Donaldson and Wollert (2008) by using Bayes s Theorem to mathematically adjust the predicted recidivism estimate from the Static-99 reference group at the obtained total score assigned to the SVP undergoing assessment.

Conclusions and Recommendations

The developers of the October 2008 Static-99 experience tables (Helmus et al., 2008) have released what they purport as risk information that is more reliable and valid than and replaces the original Static-99 risk data (Hanson & Thornton, 2000) but this assertion has yet to be examined and confirmed under the light of peer review. Based on the substantive revisions of the Static-99 risk data and rules guiding its use, the Static-99 developers have failed to adhere to psychometric standards in the development and implementation of the experience tables (AERA, 1999). Clinicians and triers of fact are left with unanswered questions as to the reliability and validity of the new experience tables in applied risk assessments with groups of sexual offenders that are similar to or differ from the Complete Sample, High Risk, and CSC. Insufficient information has been published by the Static-99 developers to support the use of the Complete Sample, CSC, or High Risk experience tables in describing the risk potential of SVP s under the risk criterion. In the haste to deploy the October 2008 Static-99 experience tables with SVP s and other sexual offenders, the Static-99 developers and users should not lose sight of the ethical obligation to practice psychology in a manner that does not harm the client by rendering inaccurate risk opinions that falsely classify SVP s as meeting the risk criterion. Toward this end, several suggestions are made for test users and the test developers.

With many unanswered questions about the reliability and validity of the October 2008 experience tables, as discussed in this paper, clinicians should have serious reservations about reporting the risk information as recommended in the Static-99 report template. The many deficiencies in the design and implementation of the October 2008 Static-99 experience tables raise questions as to whether these data should be used at all. Should clinicians decide to use this information, several recommendations are made to improve the accuracy of reporting results.

First, the risk information from October 2008 Static-99 experience tables should not be represented as normative data. Rather, evaluators should point out the lack of information provided by the Static-99 test developers such as the failure to provide empirically based interpretation rules and lack of sample representativeness, all of which compromise the confidence by which the risk data reflects an individual s potential for sexual reoffense. Second, clinicians should provide a sound justification for the selected reference group in which to compare the individual being assessed. Based on the arguments presented in this paper, clinicians would be best advised to report the risk information from the Complete Sample, as it likely provides the most reliable risk data relative to the CSC and High Risk groups. Clinicians should still qualify risk assessment opinions based on the divergence of the characteristics of the reference sample from the individual being assessed. Third,

the average logistic regression estimates and associated 95% confidence intervals should be reported for the corresponding score-wise risk level at the five and ten year follow-up periods. Reporting the risk data in this fashion provides the trier of fact relevant information necessary to assign appropriate weight in determining the fit of the Static-99 findings in meeting the statutorily defined risk criterion and deciding to what extent, if any, the selected October 2008 Static-99 experience table is representative of the risk level of the individual being assessed (Janus & Prentky, 2003), as well as conforming to ethical standards in the reporting of test results (AREA, 2003; American Psychological Association, 2000). Finally, for sexual offenders over the age of forty, clinicians may want to use the age and actuarial adjusted risk data as reported by Waggonner et al. (2008).

The Static-99 developers need to perform a more comprehensive analysis of the data for the various samples with an emphasis on determining which group, if any, is the best fit for SVP s. With the exception of the confidence intervals, the efficacy of the experience tables remains untested. The Static-99 test developers should provide data on predictive accuracy (e.g., correlation coefficients, sensitivity and specificity, and Receiver Operator Characteristics Area Under the Curve). Without these measures of reliability and validity, clinicians cannot inform triers of fact about the degree of error in risk assessment opinions necessary for them to give appropriate weight to the proffered evidence. Rules related to the application of the instrument in applied risk assessments should be empirically supported as opposed to being subjected to the vagaries of clinical judgment so as to maintain the transparency necessary to determine the extent to which the selected experience tables and the characteristics of subjects underlying this data fit with the required risk criterion. Finally, this information needs to be published and subjected to peer review.

References

1. Abracen, J., & Looman, J. (2006). Evaluation of civil commitment criteria in a high risk sample of sexual offenders. Journal of sexual offender civil commitment: Science and the law, 1, 124-139.
2. Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric assessment of dynamic risk factors for child molesters. Sexual Abuse: A Journal of Research and Treatment, 19, 347-367.
3. American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2003). Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association
4. American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. Washington, D.C.: American Psychological Association.
5. Andrews, D. A., & Bonta, J. S. (2006). The psychology of criminal conduct (4th Ed.). Cincinnati, OH: Anderson.
6. Anastasi, A., & Urbina, S. (1997). Psychological testing (7th ed.). Upper Saddle River, NJ; Prentice Hall.
7. Arizona Department of Correction (1998). Fact Sheet 98-06: Sex Offender Recidivism. Downloaded on September 14, 2008 from: http://www.adcprisoninfo.az.gov/FACTSHEETS /Fact%20Sheet%2098-06.htm.
8. Barbaree, H. E., & Blanchard, R. (2008). Sexual deviancy over the lifespan. In D. R. Laws & W. T. O Donohue (Eds.), Sexual deviance: Theory, Assessment, and Treatment (pp. 37-60), New York: Guildford Press.
9. Barbaree, H. E., Langton, C. M., & Blanchard, R. (2007). Predicting recidivism in sex offenders using the VRAG and SORAG: The contribution of age at release. International Journal of Forensic Mental Health, 6(1), 29-46.

10. Barbaree, H. E., Langton, C. M. Blanchard, R., & Cantor, J. M. (2008). Aging versus stable enduring traits as explanatory constructs in sex offender recidivism: Partitioning actuarial prediction into conceptually meaningful components. Criminal Justice and Behavior, 36(5), 443-465.
11. Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (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.
12. Brennan, R. L. (2001). Generalizability Theory. New York: Springer-Verlag.
13. California Sexual Offender Management Board (2008a). An assessment of current management practices of adult sex offenders in California: Initial report (January 2008). Sacramento, CA: Sexual Offender Management Board.
14. California Sexual Offender Management Board (2008b). Recidivism study of paroled sex offenders: A five year study. Unpublished statistics. Available from author upon request.
15. California Sexual Offender Management Board (2008c). Recidivism study of paroled sex offenders: A ten year study. Unpublished statistics. Available from author upon request.
16. Campbell, T. W. (2007). Assessing Sex Offenders (2nd Ed.). Springfield, IL: Charles C. Thomas Publishers.
17. Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. American Psychologist, 60(2), 170-180.
18. deVogel, V., deRuiter, C., van Beek, D., & Mead, G. (2004). Predictive validity of the SVR-20 and Static-99 in a Dutch sample of treated sex offenders. Law and Human Behavior, 28, 235-251.
19. Donaldson, T., & Wollert, R. (2008). A mathematical proof and example that Bayes s Theorem is fundamental to actuarial estimates of sexual recidivism risk. Sexual abuse: A Journal of Research and Treatment, 20(2), 206-218.
20. Doren, D. (2008). What do the new actuarial findings mean for real-life risk assessments? Workshop presented at the ATSA 27th Annual Research and Treatment Conference on October 23, 2008, Atlanta: GE. Available at static99.org.
21. Doren, D. M. (2002). Evaluating sex offenders. Thousand Oaks, CA: Sage.
22. Doren, D. (2004). Stability of the interpretative risk percentages for the RRASOR and Static-99. Sexual Abuse: A Journal of Treatment and Research, 16(1), 25-36.
23. Doren, D., & Thornton, D. (2008). New Norms for Static-99: A Briefing. A workshop sponsored by Sand Ridge Secure Treatment Center on November 10, 2008. Madison, WI. Available from author upon request.
24. Eher, R., Rettenberger, M., Schilling, F., & Pfafflin, F. (2008). Failure of Static-99 and SORAG to predict relevant reoffense categories in relevant sexual offender subtypes: A prospective study. Sexual Offender Treatment, 8(1), p. 1-20.
25. Epperson, D. L. (2003). Validation of the MnSOST-R, Static-99, and RRASOR with North Dakota Prison and Probation Samples. Unpublished Technical Assistance Report, North Dakota Division of Parole and Probation. Available from author upon request.
26. Fazel, S., Sjostedt, G., Langstrom, N., & Grann, M. (2007). Risk factors for criminal recidivism in older sexual offenders. Sexual abuse: A Journal of Research and Treatment 18(2), 159-167.
27. Furby, L., Weinrott, M. R., & Blackshaw, L. (1989). Sex offender recidivism: A review. Psychological Bulletin, 105(1), 3-30.
28. Hanson, R. K. (2008). Training conducted for the Sexual Offender Commitment Defense Association, October 25, 2008. Atlanta, GA.
29. Hanson, R. K. (2006). Does the static-99 predict recidivism among older sexual offenders? Sexual abuse: A Journal of Research and Treatment, 18, 343-355.

30. Hanson, R. K. (2002). Recidivism and age: Follow-up data from 4,673 sexual offenders. Journal of Interpersonal Violence, 17(10), 1046-1062.

31. Hanson, R. K. & Bussiere, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. Journal of Consulting and Clinical Psychology, 66(2), 348-362.

32. Hanson, R. K., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W., Quinsey, V. L., & Seto, M. C. (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. Sexual Abuse: A Journal of Research and Treatment, 14(2), 169-194.

33. 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 (2007-05). Ottawa, Canada: Public Safety Canada.

34. Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis. Psychological Assessment, 21(1), 1-21.

35. Hanson, R. K., & Morton-Bourgon, K. E. (2004). Predictors of sexual recidivism: An updated meta-analysis. Canada: Dept of Solicitor General.

36. Hanson, R. K., & Thornton, D. (2000). Improving risk assessment for sex offenders: A comparison of three actuarial scales. Law and Human Behavior, 24(1), 119-136.

37. Hare, R. D. (2003). Hare PCL-R (2nd Ed.). North Tonawanda, NY: Multi Health Systems.

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

39. Harris, A. J. R., & Hanson, R. K. (2004). Sexual offender recidivism: A simple question. Canada: Dept of Solicitor General.

40. Harris, A. J. R., Hanson, K., & Helmus, L. (2008). Are new norms needed for Static-99? Workshop presented at the ATSA 27th Annual Research and Treatment Conference on October 23, 2008, Atlanta: GA. Available at www.static99.org.

41. Harris, A. J. R., Phenix, A., Hanson, R. K., & Thornton, D. (2003). Static-99 Coding Rules Revised- 2003. Canada: Dept of Solicitor General.

42. 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.

43. Helmus, L. (2007a). A multi-site comparison of the validity and utility of the static-99 and static-2002 for risk assessment with sexual offenders. Unpublished honors thesis. Ottawa, Ontario, Canada: Carleton University.

44. Helmus, L. (2008a). Static-99 Recidivism percentages by risk level: Revised January 2008. Unpublished supplementary data. Available upon request from author.

45. Helmus, L. (2008b). Personal communication.

46. Helmus, L., Hanson, R. K., & Thornton, D. (2009). Reporting Static-99 in light of new research on recidivism norms. The Forum, 21(1), Winter 2009, 38-45.

47. Iowa Department of Human Rights. (2000). The Iowa sex offender registry and recidivism. Available at: www.iowa.gov/dhr/cjjp/images/pdf/01_pub/SexOffenderReport.pdf.

48. Institute of Public Policy (2006). Sex offender recidivism in Missouri and community correction options. Columbia, MO: University of Missouri, Truman School of Public Affairs.

49. Janus, E. S. (2006). Failure to protect: America s sexual predator laws and the rise of the preventive state. Ithaca, NY: Cornell University Press.

50. Janus, E. S., & Prentky, R. A. (2003). Forensic use of actuarial risk assessment with sex offenders: Accuracy, admissibility, and accountability. American Criminal Law Review, 40(1143), 1-59.

51. Kalton, G. (1983), Introduction to Survey Sampling, SAGE University Paper series on Quantitative Applications in the Social Sciences, series no. 07-035. Beverly Hills and London: SAGE Publications, Inc.

52. Kazdin, A. E. (2003). Research design in clinical psychology. Boston, MA: Allyn & Bacon.
53. 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). Submitted to the U.S. Department of Justice.
54. Langan, P. A., Schmitt, E. L., & Durose, M. R. (2003). Recidivism of sex offenders released from prison in 1994. Washington, D.C.: U.S. Dept of Justice.
55. Langton, C. M. (2003). Contrasting approaches to risk assessment with adult male sexual offenders: An evaluation of recidivism prediction schemes and the utility of supplementary clinical information for enhancing predictive accuracy. Dissertations Abstracts International, 64
(04), 1907B. (UMI No. NQ78052).
56. Långström, N. (2004). Accuracy of actuarial procedures for assessment of sexual offender recidivism risk may vary across ethnicity. Sexual Abuse: A Journal of Research and Treatment, 16, 107-120.
57. Långström, N., Sjostedt, G., & Grann, M. (2004). Psychiatric disorders and recidivism in sexual offenders. Sexual abuse: A Journal of Research and Treatment, 16(2), 139-150.
58. Looman, J. (2006). Comparison of two risk assessment instruments for sexual offenders. Sexual abuse: A Journal of Research and Treatment, 18(2), 193-206.
59. Losel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. Journal of Experimental Criminology, 1, 117-146.
60. Marques, J. K., Weideranders, M., Day, D. M., Nelson, C., & van Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from the California s sex offender treatment and evaluation program (SOTEP). Sexual Abuse: A Journal of Research and Treatment, 17(1), 79-107.
61. Miller, H. A., Amenta, A., & Conroy, M. (2005). Sexually violent predator evaluations: Empirical evidence, strategies for professionals, and research directions. Law and Human Behavior, 29, 29 54.
62. Milloy, C. (2007). Six year follow up of 135 released sex offenders recommended for commitment under Washington s sexually violent predator law (June). Washington State Institute for Public Policy.
63. Minnesota Department of Corrections (2007). Sexual offender recidivism in Minnesota: April 2007. Available from www.doc.state.mn.us
64. Padilla, J. (2006). Unpublished study data. October 10, 2006 memo to Jim McEntee, Public Defender. Available upon request from author.
65. People v. Ghilotti. 119 California Reporter. 2d 1. (2002).
66. Phenix, A., & Arnold, D. (2008). Proposed Considerations for Conducting Sex Offender Risk Assessment. Presentation to California Department of Mental Health (December 14, 2008). Unpublished paper. Available upon request from author.
67. Prentky, R. A., Janus, E., Barbaree, H., Schwartz, B., & Kafka, M. (2006). Sexually violent predators in the courtroom: Science on trial. Psychology, Public Policy, and Law, 12, 357 393.
68. Prentky, R. A., & Lee, A. F. S. (2007). Effect of age-at-release on long term sexual re-offense rates in civilly committed sexual offenders. Sexual Abuse: A Journal of Research and Treatment, 19, 43-59.
69. Saum, S. (2007). A comparison of an actuarial risk prediction measure (Static-99) and a stable
dynamic risk prediction measure (Stable-2000) in making risk predictions for a group of sexual
offenders. Dissertations Abstracts International, 68 (03), B. (UMI No. 3255539).
70. Sjostedt, G., & Grann, M. (2002). Risk assessment: What is being predicted by actuarial prediction instruments. International Journal of Forensic Mental Health, 1(2), 179-183.

71. Sjostedt, G., & Långström, N. (2001). Actuarial Assessment of Sex Offender Recidivism Risk: A cross-validation of the RRASOR and the Static-99 in Sweden. Law and Human Behavior, 25(6), 629-644.
72. Song, L., & Leib, R. (1995). Washington State Sex Offenders: Overview of Recidivism Studies (February 1995). Washington: Washington State Institute of Public Policy.
73. State of Delaware (2007). Recidivism of Delaware Adult Sex Offenders Released from Prison in 2001(July 2007). Delaware: Office of Management and Budget Statistical Analysis Center.
74. State of New York (2007). Research Bulletin: Sex Offender Populations, Recidivism and Actuarial Assessment. New York: New York State Division of Probation and Correctional Alternatives. Available at: http://www.dpca.state.ny.us/pdfs/somgmtbulletinmay2007.pdf.
75. State of Ohio, Department of Rehabilitation and Corrections (2001, April). Ten-year follow up study of 1989 sex offender releases. Ohio: Department of Rehabilitation and Corrections.
76. Stockburger, D. (1998). Introductory statistics: Concepts, models, and applications. Available at: http://www.psychstat.missouristate.edu/introbook/sbk00.htm. Last accessed December 31, 2008.
77. Tennessee Bureau of Investigations (2007). Recidivism Study. Available at: http://www.tbi.state.tn.us/Info%20Systems%20Div/TIBRS_unit/Publications/Sex%20Offender%20Re
78. Thornton, D. (2006). Age and sexual recidivism: A variable connection. Sexual Abuse: A Journal of Research and Treatment, 18(2), 123-135.
79. Tulsky, D., Zhu, J., & Ledbetter, M. E. (2002).WAIS-III & WMS III: Technical Manual. San Antonio, TX: The Psychological Corporation.
80. Waggoner, J., Wollert, R., & Cramer, E. A Re-specification Of Hanson s Updated Static-99 Experience Table That Controls For The Effects Of Age On Sexual Recidivism Among Young Offenders. Law, Probability, and Risk, 7(4), 305-312.
81. Washington State Institute for Public Policy (2005, August). Sex offender sentencing in Washington State: Measuring recidivism. 1-4.
82. Woodworth, G. G., & Kadane, J. (2004). Expert testimony supporting post-sentence civil incarceration of violent sex offenders. Law, Probability and Risk, 3, 221 241.
83. Wollert, R. W. (2006). Low base rates limit expert certainty when current actuarial tests are used to identify sexually violent predators: An application of Bayes s Theorem. Psychology, Public Policy, and Law, 12, 56 85.
84. Wollert, R. (2007a). Poor diagnostic reliability, the null-bayes logic model, and their implications for sexually violent predator evaluations. Psychology, Public Policy, and the Law, 13(3), 167 203.
85. Wollert, R. (2007b). Validation of a bayesian method for assessing sexual recidivism risk. Paper presented at the 115th annual convention of the American Psychological Association, San Francisco, CA.
86. Wyoming Legislative Service Office. (2005). Research memo. Available at: www.legisweb.state.wy.us

Author's note

Notes

[1]The type of sexual reoffending (e.g., sexually violent offenses or any type of sexual reoffense) to be predicted by the risk criterion may vary among jurisdictions; however, this does not affect the

threshold of risk that must be met. California law (People v. Ghilotti, 2002) specifically prohibits evaluators from assigning a specific proportion of risk such as 30% or 50% and allows a positive finding on the risk criterion at less than a chance level of engaging in future sexually violent predatory behavior, as long as it is considered a high risk.

[2]The October 2008 reporting procedures are available upon request from the author.

[3]The methodological deficiencies in developing the current selection criteria likely contribute to the unstable confidence intervals found in the CSC and High Risk groups since the reliability of the scoring system has an inverse relationship on the width of the confidence intervals (i.e., lower reliability of the scoring corresponds with wider confidence intervals).

[4]Available at: http://www.static99.org/pdfdocs/percentilestablejanuary2009.pdf

[5]Available at: http://www.static99.org/pdfdocs/standardreportingparagraphjanuary2009.pdf

Author address

Brian R. Abbott, Ph.D.
Forensic Psychological Services
111 N. Market Street, Suite 300
San Jose, CA. 95136
Phone: (408) 451-8465
brian@dr-abbott.net