Throwing the Baby Out With the Bath Water: Is It Time for Clinical Judgment to Supplement Actuarial Risk Assessment?

Brian R. Abbott, PhD

The assessment of the potential for sexual violence is one of three prongs that must be met to satisfy the requirements for civil confinement of dangerous sex offenders in the 21 U.S. jurisdictions that have these laws. In a recent issue of *The Journal*, Sreenivasan et al. argued that, because of a host of methodological problems, actuarial risk assessment methods in general and the Static-99 and its progeny in particular are insufficient for accurate assessment of risk for dangerous sex offenders. They propose using a combination of clinical judgment with actuarial science as a solution. This analysis and review of Sreenivasan et al. reveals and corrects flaws in the arguments they employed to support their position and shows how the combination of actuarial science with clinical judgment is more error prone than the actuarial approach only, and cannot be forensically defended in court. Recommendations on reporting Static-99R data in expert testimony are provided, taking into account the limitations of the instrument.

In an article recently published in *The Journal*, Sreenivasan and colleagues[1] wrote about the limitations of sexual recidivism actuarial risk instruments in assessing the dangerousness criterion under sexually violent predator or sexually dangerous person (SVP/SDP) civil confinement laws, in particular focusing on the Static-99 and Static-99R. They discussed some important points about the limitations of the Static-99 and its current revision, the Static-99R, when deployed in forensic practice. Based on these limitations, they argued that other risk-relevant clinical factors should be considered along with the results of actuarial risk assessment tools, in rendering a comprehensive risk assessment. This commentary will show that the very arguments Sreenivasan et al. propounded for combining actuarial science with clinical judgment actually lead to different conclusions.

The first section of the commentary presents various arguments of Sreenivasan et al. that are in fact misrepresentations of the literature related to the Static-99 and Static-99R. The response then addresses the legitimate arguments that Sreenivasan et al. present but shows how the limitations they assert as supporting the use of a combination of actuarial science and clinical judgment are in fact the major strengths of actuarial instruments that provide guidance to clinicians in their proper application and interpretation. The commentary concludes by discussing some suggestions for the application of the Static-99R and the possible role of empirically validated methods of combining actuarial risk assessment with other risk-relevant factors external to the actuarial instrument.

Misrepresentations

Sreenivasan et al.[1] argue that the score-wise recidivism estimates obtained with the three versions of the Static-99[2–4] and the 2006 age data published by Hanson[5] could be applied to individuals being assessed under SVP/SDP statutes. There is no justification for a forensic evaluator to apply the wide array of risk estimates, as suggested by Sreenivasan et al.

Dr. Abbott is in private practice in San Jose, CA. Address correspondence to Brian R. Abbott, PhD, 111 N. Market Street, Suite 300, San Jose, CA 95113. E-mail: brian@dr-abbott.net.

Disclosures of financial or other potential conflicts of interest: None.

With the unveiling of each successive iteration of the Static-99, including the revised version (Static-99R), the developers instructed end users to employ the revised sexual recidivism experience tables over the previous version.[3,4] The Static-99R developers have provided age-graded, score-wise risk estimates that can be used[6] in place of the 2006 age-adjusted risk estimates. The age of the individual who is being assessed should be considered in the clinician's decision to use the 5- or 10-year recidivism estimates. Published literature documenting the risk-mitigating effect of advancing age[7] and the age analysis conducted by the developers of the Static-99R[8] suggest that a reasonable rule of thumb would be to use the 5-year recidivism rate for SVPs/SDPs who are age 40 or older. The 10-year risk Static-99R risk estimates would overpredict sexual recidivism risk for individuals age 40 or older, as these data reflect rising recidivism risk in an age group showing declining risk. Given the importance of advancing age as a risk-mitigating factor,[9,10] clinicians should report and testify about the Static-99R age-adjusted and unadjusted recidivism rates and make it clear to triers of fact that actuarial risk estimates do not reflect the recidivism risk of the individual who is being assessed.[11]

A second area of criticism leveled by Sreenivasan *et al.* is that the experience tables contained in the second and third iterations of the Static-99 (2008 and Static-99R) may be defective because the developers relied on unpublished studies. This possibility would be a legitimate concern if the Static-99 developers had relied on data contained only in the unpublished papers. They did not, however. They acquired and reanalyzed the raw data from each of the published and unpublished studies, from which they calculated the sexual recidivism data provided in the Static-99 and Static-99R experience tables.[12] Since it is rare for peer reviewers to examine raw data sets when considering manuscripts submitted for publication, the criticism about the use of unpublished papers lacks merit. On the other hand, it is incumbent on the Static-99R developers to report clearly on the methodology leading to the development and implementation of the Static-99R experience tables in a manuscript that is subject to peer review and publication.

Sreenivasan *et al.* make valid remarks about sample representativeness, which will be discussed later. Briefly, however, they make two misrepresentations regarding the Static-99R Preselected High Risk Group. They describe this group as comprising sex offenders who were referred to forensic psychiatric hospitals and characterize one of the samples from Bridgewater, Massachusetts, as consisting of insanity acquittees. Both characterizations are incorrect. The Preselected High Risk Group comprises sex offenders from high-security forensic hospitals and prisons in Canada.[10] The Bridgewater sample comprises sex offenders who were sent to that facility for consideration as sexually dangerous persons (SDPs)[13] under a version of the civil confinement law in Massachusetts that was repealed in 1990.[14] An unknown proportion were committed as SDPs, while the remaining offenders were returned to prison to serve their criminal sentences. This distinction in the Bridgewater sample is important because those offenders who were found not to be SDPs and who were returned to prison to complete their sentences may be more appropriately classified in one of the other Static-99R reference groups (e.g., Routine Corrections or Preselected Treatment). Such criterion contamination diminishes the validity of the risk data from the Preselected High Risk Group when trying to apply it to other groups of sex offenders, including SVP/SDP populations.

Sreenivasan *et al.* discuss the differences in calculating sexual recidivism estimates by logistic regression versus survival analysis. They seem to imply that the Static-99R developers' instructions to use the logistic regression estimates may be faulty because the survival analysis method is superior, although they fail to illustrate whether this is actually true when applied to the Static-99 and Static-99R. Rather than get into a lengthy discussion about the mechanics of logistic regression analysis and survival analysis, it is easier to test the assertion of Sreenivasan *et al.* by using both methods of calculation to examine the Static-99R experience tables that report the recidivism rates. Inspection of these experience tables[14] reveals the fallacy of the authors' argument. Each score-wise logistic regression estimate contained in the experience tables includes a 95 percent confidence interval. If a score-wise risk estimate determined by survival analysis falls outside the 95 percent confidence interval for the corresponding logistic regression estimate, then it might be concluded that the difference between the two methods of calculation is meaningful. With rare exception, the score-wise risk estimates based on survival analysis fall within the 95 percent confidence interval for the lo-

gistic regression estimates, and the two estimates therefore do not differ significantly. The Preselected High Risk Group score-wise risk estimates at the 10-year follow-up reflect the least congruence between the two statistical methods. In this situation, the logistic regression data reflect higher sexual recidivism estimates than do the results of the survival analyses.

Finally, Sreenivasan *et al.* fault the developers of the Static-99 and Static-99R for implying that the score-wise risk estimates were stable across populations, when in fact they were not. The basis of this criticism seems to lie in a misunderstanding of the metric used to assess predictive accuracy. The authors appear to conflate the effect size known as the receiver operating characteristics area under the curve (AUC) as a metric to assess the stability of score-wise risk estimates across populations. The AUC is a rough measure of predictive accuracy that indicates the extent to which the Static-99 or Static-99R separates recidivists from nonrecidivists at each score. The AUC is not a measure of predictive accuracy of risk estimates but rather plots the true- and false-positive rates at each score on the actuarial measure. Moreover, the AUC found for a specific sample to which the Static-99 or Static-99R has been applied is usually not generalizable to other populations.[15]

Clinicians familiar with the peer-reviewed published literature became aware of the instability of the Static-99 score-wise risk estimates as early as 2004, when Doren[16] published data comparing Static-99 risk estimates across populations. I discussed the variability in Static-99 score-wise risk estimates across populations in a paper that was published in 2009.[17] Also in 2006, Mossman[18] described how sexual recidivism score-wise risk estimates are determined entirely by the base rate of sexual recidivism and the distribution of score-wise likelihood ratios by the actuarial instrument in a particular population of sex offenders. When applying these principles to Doren's results, Mossman found greater instability in the Static-99 score-wise risk estimates across samples than Doren reported. Mossman's work was validated in a mathematical proof published by Donaldson and Wollert[19] in 2008. As cited by Sreenivasan *et al.*, there was a significant lapse of time before the Static-99 experience tables reflected reductions in recidivism risk as a function of lower base rates. This circumstance, however, does not relieve the end users who employ the Static-99 of the responsibility of

considering and applying currently available base rate information when interpreting Static-99 score-wise risk estimates that are based on substantially different base rates than those found in local populations of sex offenders.[20] This topic is discussed in the conclusion section.

Analysis of the Legitimate Criticisms

In the end, Sreenivasan *et al.* cite two legitimate criticisms about the methods used to develop and validate the Static-99 and Static-99R: sample representativeness and uniform measurement of outcome. This section briefly comments on these two arguments and then addresses whether these deficiencies justify combining clinical judgment with actuarial science.

As Sreenivasan *et al.* note, the Static-99 (2008 version) and the Static-99R combine recidivism data representing arrests and convictions. This artifact is a result of the various raw data sets that the Static-99/-99R developers collected when the original researchers decided on the definition of the recidivism criterion. Using data provided by Helmus (Table D-1; Ref. 12, p 160), I calculated that 60 percent of the recidivism estimates are based on convictions for sexual offenses and the remaining proportion on charges. Clearly, this result presents a problem related to the interpretation of the risk estimates and raises the question of the extent to which data based on convictions may underestimate the true proportion of sexual recidivism. It is generally accepted in the field that recidivism data based on convictions underestimate the true sexual recidivism rate. Solving this problem and arriving at a way to estimate the true recidivism rate is complicated and beyond the scope of this response. Interested readers are referred to Wollert[21] for a detailed discussion. Nonetheless, this deficiency does not justify combining actuarial science with clinical judgment, nor would this method solve this problem. Rather, clinicians must state in written reports and oral testimony the limitations of using Static-99R risk estimates, in that they are based on a combination of charges and convictions.

Sample representativeness is a serious concern with the Static-99 and Static-99R. I authored a critical analysis of the 2008 version of the Static-99,[17] which, among many topics, explained how the lack of sample representativeness was a significant threat to the validity of the Static-99 risk estimates when

applied to other groups representing individuals who were being assessed. In the analysis of the data underlying the Static-99R, Helmus[12] reported that she took into account many of the criticisms that I raised regarding sample representativeness. This exchange in part led the Static-99R developers to employ data analysis methods to control for threats related to sample representativeness that ultimately resulted in the conception of the four-reference-group model (Routine Corrections, Nonroutine Corrections, Preselected Treatment, and Preselected High Risk).

This model, used for the interpretation of Static-99R score-wise risk estimates, is of questionable validity and reliability, but for reasons different from those pointed out by Sreenivasan *et al.* The model was devised by a combination of statistical discrimination and clinical judgment.[12] Three of the four reference groups (Routine Corrections, Preselected Treatment, and Preselected High Risk) were each found to be significantly different from one another on the variable of sexual recidivism base rates. The Nonroutine Corrections Group was constructed by combining the Preselected Treatment and High Risk groups with three additional samples, and the base rate was not significantly different from those of the other three groups.[12] The Static-99R developers employed *ad hoc* logical inference to develop criteria they believe distinguish the four reference groups so that clinicians in the field could select an appropriate reference group to compare with the group representing the individual who is being assessed. This procedure assumes that the criteria distinguishing the Static-99R reference groups are valid and can be reliably applied to other groups of sex offenders. The Static-99R developers provide no validity or reliability data that this method works. It is unlikely that this will ever be accomplished because the data necessary to test the discrimination criteria for each reference group were not collected on each sample comprising the Static-99R data set. The only empirically validated method to select a reference group is based on the similarity of base rates between the reference group and the group representing the individual who is being assessed.

Sreenivasan *et al.* raise an important question related to sample representativeness, regarding the difference between individuals in a comparison group and those in the Static-99R reference group from which the risk estimates were derived. In practice, score-wise risk estimates are interpreted by comparing the extent of similarity between the selected Static-99R reference group and the group representing the individual who is under assessment. There is a reciprocal association in this relationship, in that the fewer distinguishing characteristics shared between the two groups, the less valid the score-wise risk estimates from the Static-99R reference group when applied to the group representing the individual who is being evaluated. Needless to say, there is no easy way to quantify this difference and to adjust accordingly the probability of risk provided in the actuarial experience tables. Even if this problem could be solved, there is an additional dilemma as to how group-based, score-wise risk estimates are applied to the individual who is under assessment.

Cooke and Michie[22] offer a solution to this predicament by applying a mathematical formula, expressed by what is known as the confidence interval for the individual (CII), that accounts for the extent of error in the group-based, score-wise risk estimates when applied to an individual. The CII is conceptually similar to a 95 percent confidence interval, but it accounts for the standard error of estimation in prediction when the score-wise risk estimate is applied to the individual. The CII depends on the magnitude of the correlation between the risk instrument and sexual recidivism as applied to the individual. The lower this correlation, the higher the degree of error in the score-wise risk estimate as applied to the individual. In any group of sex offenders, the actuarial instrument may be highly associated with the potential to engage in future sexual recidivism, while the extent of this association is lower or possibly absent for others in the population. The CII quantifies the variance for individuals in the population in the form of a level of confidence whereby a clinician can determine the extent to which he can rely on the score-wise risk estimate as applying to the individual whom he is assessing. This information can be communicated in expert testimony so that the trier of fact can assign appropriate weight to the risk assessment outcome, as statutorily defined in the SVP/SDP laws. Interested readers are referred to Donaldson and Abbott[23] for a further explanation as to how the CII is calculated with dichotomous outcome data, such as sexual recidivism and its application to the Static-99R.

The CII can be used to demonstrate conceptually that actuarial assessment combined with clinical judgment increases the extent of error over that of the

actuarial-only method when group-based, score-wise risk estimates are applied to the individual. As described in the previous paragraph, the standard error of estimation used in the CII is based on the correlation between the predictor and outcome variables. The correlation between the Static-99 and the prediction of sexual recidivism ranges between 0.245[24] and 0.33,[2] while clinical judgment correlates at a rate of 0.10 with the same outcome variable.[25] Given that clinical judgment correlates at a lower rate than the Static-99, it would be expected that the combination of the two methods would yield a correlation coefficient somewhere between their respective individual values. This low correlation translates into an enormous amount of error variance, wherein the score-wise risk estimates would be likely to span the entire range of predicted values (i.e., 0%–100%), assuming that it is even possible to derive score-wise estimates using the method proposed by Sreenivasan *et al.*

In Weakness, There Is Strength

The methods by which actuarial instruments are constructed bring transparency, accountability, and consistency to the judicial risk-finding process in SVP/SDP proceedings that is not possible when using clinical judgment.[11] As weaknesses are identified in the development of actuarial risk methods (e.g., changing base rates and threats to external validity of risk estimates), the instrument developers can study ways to correct such deficiencies and distribute solutions to end users, as has been the situation with the developers of the Static-99 and Static-99R. By understanding the limitations of the methods by which actuarial instruments are developed, clinicians can more accurately testify about the relevancy or admissibility of measures like the Static-99R. For instance, in a recent *Daubert* challenge to the Static-99R in a New Hampshire SVP/SDP case, the judge found most of the Static-99R inadmissible except for the data from the Routine Corrections Group, largely because of the lack of evidence regarding the reliability and validity used to develop the four-reference-group model.[26] Taking this example one step further, experts who now testify about the Static-99R in New Hampshire can explain the strengths and limitations of the risk information related to the Routine Corrections Group, so that the trier of fact can assign appropriate weight to the risk data in determining whether the individual who is being assessed meets the statutorily defined risk criterion.

Clinical Judgment Is a Weaker Approach

Sreenivasan *et al.* cite one study indicating that the combination of Static-99 with clinical judgment did not improve the accuracy of predicting sexual recidivism.[27] They appear to dismiss the findings of the study by stating that the results were obtained because probation officers made the ratings, which may be valid but misses a critical point regarding the robust body of literature that consistently shows that actuarial methods of prediction far surpass clinical judgment in accuracy. As early as 1956, Meehl[28] discussed research showing the superiority of actuarial assessment methods over clinical judgment unrelated to sex offenders and sexual recidivism. Grove and Meehl[29] summarized a large body of literature (starting in the 1960s) supporting actuarial science over clinical judgment, as well as discussing the fallacy of clinicians who attempt to combine the two approaches, as advocated by Sreenivasan *et al.* Literature that specifically addresses actuarial versus clinical judgment in predicting sexual recidivism risk has been published since 1998. There have been two major meta-analyses that have examined the use of clinical judgment versus actuarial methods in assessing sexual recidivism risk.

In 1998, Hanson and Bussiere[25] conducted a meta-analysis of 10 studies using clinical judgment in an aggregate sample of 1,453 sex offenders and found that the correlation between this method of assessment and the prediction of sexual recidivism was poor. The correlation between actuarial methods and predicting future sexual recidivism ($n = 6$ studies; aggregate sample size = 684) was about four times higher than clinical judgment. Hanson and Morton-Bourgon[30] updated the 1998 meta-analysis by examining nine studies using clinical judgment, where five of the studies were contained in the 1998 meta-analysis, with a total aggregate sample size of 1,679. They reported aggregate data from 33 sexual recidivism actuarial risk assessment studies with a combined sample size of 6,972. The median and mean effect sizes (Cohen's d) for the actuarial methods were almost one-half and one-third greater, respectively, than those for clinical judgment. The peer-reviewed literature overwhelmingly rebuts the contention made by Sreenivasan *et al.* that the combination of actuarial science with clinical judgment improves predictive accuracy.

Combining actuarial science with clinical judgment obscures the transparency, accountability, and

consistency necessary to establish the reliability and validity of the risk prediction, to ensure that it meets standards of relevancy, admissibility, and weight. Sreenivasan *et al.* do not offer any reference in defense of their position that combining actuarial science with clinical judgment has been empirically supported or conceptually endorsed.

In the conclusion to their article, Sreenivasan *et al.* refer to the importance of considering salient clinical risk factors in conjunction with actuarial risk assessment results when practitioners offer risk opinions. As examples of this idea, they identify the risk factors of strangulation and sexual sadism. They offer no references to support that these two risk factors increase the potential to engage in sexual reoffending let alone incrementally increasing risk beyond that determined by the Static-99R. This writer is not aware of any sexual recidivism studies that state that acts of strangulation involving multiple victims increase the risk of sexual recidivism. The closest proxy variable to strangulation as studied in meta-analyses may be paraphilic interest in rape or violence, the degree of force used,[25] or the force or injury to victim,[30] all of which were found not to be significantly related to predicting the occurrence of sexual recidivism. Kingston *et al.*[31] directly addressed the external risk factor of sexual sadism and found that it was unrelated to predicting future sexual recidivism. These findings illustrate a significant deficiency in the risk assessment approach propounded by Sreenivasan *et al.*

Sreenivasan *et al.* assert that the combination of actuarial science with clinical judgment somehow improves the accuracy of risk assessment in SVP/SDP proceedings. The abject lack of literature to justify their approach compared with the peer-reviewed literature that contradicts the validity of the approach illustrates the danger of the method that they propose. Clinical judgment relies on the idiosyncratic opinions of clinicians based on risk factors that the clinicians believe increase the potential for sexual recidivism without having to prove empirically that such relationships exist. Clinicians who combine actuarial science with clinical judgment cannot inform the trier of fact regarding the validity, reliability, or error rate of the method. Most importantly, there is no method by which clinicians can adjust the actuarially derived score-wise risk estimate to account for the purported risk-enhancing effect of

other salient clinical factors.[1] Returning to the case example in Sreenivasan *et al.* illustrates this last point.

Their example patient (whom they called Mr. X) had a score of four on the Static-99R. The five-year recidivism estimate for those with this score in the Preselected High Risk Group is 20.1 percent (assuming that this comparison group can be justified as applicable to the group representing the patient). If he is being assessed for civil confinement as an SVP/SDP where the legally prescribed level of risk is greater than 50 percent, it becomes apparent that sex offenders comprising the group of which he is a member would not meet this threshold based on the findings from the Static-99R. The clinician then decides to incorporate what he believes are other clinically relevant risk factors, including strangulation of victims and a diagnosis of sexual sadism. By combining the actuarial approach with other salient clinical factors, the clinician testifies that the patient meets the legally prescribed burden of risk, in that there is greater than a 50 percent probability that he will engage in future sexual violence. This example raises several significant questions. What was the method used by the evaluator to increase the patient's risk potential by two and one-half times over the actuarially derived risk estimate? What studies have examined these two risk factors in relation to the probability of committing future sexual offenses? Can the probability of risk yielded by the two other clinical factors be combined with the actuarially derived risk estimate to elevate the respondent's risk by 30 percentage points? What are the reliability and validity of this method? What is the false-positive rate associated with this method? Is this approach generally accepted in the scientific community? None of these important questions related to the transparency, accountability, and consistency of the risk assessment method can be answered to justify such testimony as relevant or admissible.

Conclusions

Sreenivasan *et al.* argue that the reasoned judgments of forensic experts based on their education, skill, and professional experience when combined with actuarial science somehow improves the validity of risk assessments conducted in the SVP/SDP arena. The logic they employ to justify their position is internally inconsistent, in that they fail to apply the same standards of evaluating the efficacy of the actuarial approach to the combination of actuarial sci-

ence and clinical judgment. Rather, they assume it works to provide incremental predictive validity beyond the actuarial risk score despite offering no empirical support for this position. This commentary has provided ample evidence that the prevailing literature for more than 50 years has not supported the methods advocated by Sreenivasan *et al.* The hazard of using their approach can be seen when examining the legal requirements for civil confinement of SVPs/SDPs.

SVP/SDP laws adopt a framework for identifying convicted sex offenders who are nearing completion of their criminal sentences and authorize civilly committing them in secure treatment settings if they meet certain criteria. Although there are some variations among jurisdictions, SVP laws share common elements to justify civil confinement of individuals,[32] including the presence of a mental disorder or abnormality that causes or is associated with an elevated risk of future sexual misconduct and the purpose of commitment, which is treatment. Offenders who are legally determined to be sexually violent predators remain involuntarily confined until they no longer meet the criteria.[11] The legally defined risk threshold is commonly, but not in all jurisdictions, defined as more likely than not,[31] which can be conceptually quantified as a risk greater than 50 percent. To inform the trier of fact whether an individual meets the legally prescribed level of risk in SVP/SDP proceedings, clinicians must know the absolute score-wise risk estimates,[12,21] along with the associated degree of measurement error. Actuarial experience tables and the CII provide this information. When combining actuarially derived risk estimates with clinical judgment the score-wise risk estimates become invalid, and there is no known error rate associated with the prediction. As discussed previously, even if this problem could be solved, the resulting low correlation of this method would result in enormous error variance in the predicted group-based score-wise risk estimates that are applied to the individual. As a result, clinicians using the method advocated by Sreenivasan *et al.* would be unable to substantiate forensically that an individual who is being considered for civil confinement meets the statutorily defined risk criteria.

The combination of clinical judgment with actuarial science opens the door to confirmatory and anchoring biases that can influence evaluators' opinions as to whether individuals meet the risk criterion set forth in SVP/SDP statutes.[33] This approach encourages the clinician who is inclined to believe that the respondent meets the legally prescribed threshold of risk to conclude that he in fact does, when the actuarially derived risk estimates indicate the contrary. These same biases may operate powerfully in SVP/SDP trials. Jurors who are inclined to view a respondent as meeting the risk criterion are likely to dismiss conclusions to the contrary that are based on actuarial science in favor of reasoned judgments that resonate with the jurors' initial negative opinions of the respondent that are engendered by prosecutors who present the lurid minutia of the sexual offenses committed by the accused. The only way to counteract such partisan influences is for courts to allow only testimony about risk factors external to actuarial instruments that have been empirically supported through cross-validation in SVP/SDP populations as increasing predictive validity beyond the actuarial score-wise risk estimate.

Research in this direction holds some promise,[33–36] but is only in its infancy and requires cross-validation in other groups, including SVP/SDP populations. On the other hand, studies examining the predictive validity of the Static-99 combined with other dynamic factors assessed by a structured assessment instrument found that this method fell below chance levels of prediction in sex offenders from the United States[36] and either decreased or did not improve the detection rate of sexual recidivists beyond the Static-99 alone.[37,38] Given the current state of the literature, clinicians should refrain from considering and courts should disallow any testimony about the use of clinical risk factors outside the actuarial instrument to arrive at a conclusion about whether respondents meet the legally prescribed SVP/SDP threshold of risk for sexual violence.

At this time, the scientific literature supports the use of an actuarial-only approach in assessing the sexual recidivism risk of individuals who are being considered for civil confinement as SVPs/SDPs. The Static-99 has the greatest amount of published research to support its use compared with other sexual recidivism actuarial instruments. It is also the most frequently employed actuarial instrument in SVP/SDP assessments.[39] A major limitation in using the Static-99R is the lack of peer-reviewed literature about the method. Despite the deficit in published studies, it is reasonable to conclude that the Static-99R has validity and reliability similar to that of its

predecessor instrument based on the empirical methods used to revise the scoring of the age-at-release item.[8] Based on the known limitations of the Static-99R as documented in this response and elsewhere (Abbott BR, unpublished paper, 2010), the following three recommendations are offered for utilizing it in SVP/SDP risk assessments.

First, clinicians should not employ the discrimination criteria offered by the Static-99R developers to select a reference group to compare with the population representing the individual who is being assessed. Instead, practitioners should select the Static-99R reference group that best matches the known local base rate for prison releasees or SVPs/SDPs in the local community. In the absence of local base rate data, clinicians can rely on national base rate information[40] of 6.5 percent over five years, which would support using the score-wise risk estimates from the Routine Corrections Group.

Second, the age-unadjusted and -adjusted risk estimates should be reported. For those offenders aged 40 or older, the five-year recidivism rates should be presented with the caveat that the risk of sexual recidivism with each year of advancing age declines at a rate of between two[8] and four[10] percent.

Third, based on the reference group selected, clinicians can apply the CII for the individual's score-wise risk estimate, which becomes the basis for informing the trier of fact as to the level of confidence that can be assigned to the score-wise risk estimate. Donaldson and Abbott[23] discuss this procedure in detail and provide the CII's for each of the Static-99R reference groups.

References

1. Sreenivasan S, Weinberger LE, Frances A, *et al*: Alice in actuarial land: through the looking glass of changing Static-99 norms. J Am Acad Psychiatry Law 38:400–6, 2010
2. Hanson RK, Thornton D: Improving risk assessment for sex offenders: a comparison of three actuarial scales. Law Hum Behav 24:119–36, 2000
3. Helmus L, Hanson RK, Thornton D: Reporting Static-99 in light of new research on recidivism norms. Forum 21:38–45, 2009
4. Hanson RK, Phenix A, Helmus L: Static-99 and Static-2002: how to interpret and report scores in light of recent research. Workshop presented at the ATSA (Association for the Treatment of Sexual Abusers) 28th Annual Research and Treatment Conference, Dallas, TX, September 2009
5. Hanson RK: Does the Static-99 predict recidivism among older sexual offenders? Sex Abuse J Res Treat 18:343–55, 2006
6. Helmus L, Thornton D: Static risk assessment: understanding and applying the 2009 norms. Workshop presented at the California Department of Mental Health Sex Offender Commitment Program, Sacramento, December 2009
7. Barbaree HE: The effects of aging on sexual offender recidivism. Workshop presented at the California Department of Mental Health Sex Offender Commitment Program, Monterey, March 2010
8. Helmus L, Babchishin KM, Hanson RK, *et al*: Static-99R: revised age weights. Available at http://www.static99.org/pdfdocs/static-99randage20091005.pdf. Accessed on January 10, 2009
9. Barbaree HE, Blanchard R: Sexual deviancy over the lifespan, in Sexual Deviance: Theory, Assessment, and Treatment. Edited by Laws DR, O'Donohue WT. New York: Guildford Press, 2008, pp 37–60
10. Barbaree HE, Langton C, Blanchard R, *et al*: Aging versus stable enduring traits as explanatory constructs in sex offender recidivism: partitioning actuarial prediction into conceptually meaningful components. Crim Just Behav 36:443–65, 2009
11. Janus ES, Prentky RA: Forensic use of actuarial risk assessment with sex offenders: accuracy, admissibility, and accountability. Am Crim L Rev 40:1–59, 2003
12. Helmus L: Re-norming Static-99 recidivism estimates: exploring base rate variability across sex offender samples. Unpublished master's thesis, Ottawa, ON, Canada: Carleton University, 2009
13. Knight RA, Thornton D: Evaluating and improving risk assessment schemes for sexual recidivism: a long-term follow-up of convicted sexual offenders. Unpublished document submitted as a final report to the U.S. Department of Justice, Document 217618, March 2007
14. Detailed recidivism tables for Static-99R. Available at http://www.static99.org/pdfdocs/detailed_recid_tables_static99r_2009-11-15.pdf. Accessed November 23, 2009
15. Vrieze SL, Grove WM: Predicting sex offenders' recidivism: I. correcting for item overselection and accuracy overestimation in scale development. II. sampling error-induced attenuation of predictive validity over base rate information. Law Hum Behav 32:266–78, 2008
16. Doren D: Stability of the interpretative risk percentages for the RRASOR and Static-99. Sex Abuse J Res Treat 16:25–36, 2004
17. Abbott BR: Applicability of the new Static-99 experience tables in sexually violent predator risk assessments. Sex Offend Treat 1:1–24, 2009
18. Mossman D: Another look at interpreting risk categories. Sex Abuse J Res Treat 18:41–63, 2006
19. Donaldson T, Wollert R: A mathematical proof and example that Bayes's Theorem is fundamental to actuarial estimates of sexual recidivism risk. Sex Abuse J Res Treat 20:206–17, 2008
20. American Educational Research Association, American Psychological Association and National Council on Measurement in Education: Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association, 2003
21. Wollert R: Low base rates limit expert certainty when current actuarials are used to identify sexually violent predators. Psychol Public Pol Law 12:56–85, 2007
22. Cooke DJ, Michie C: Limitations of diagnostic precision and predictive utility in the individual case: a challenge for forensic practice. Law Hum Behav 34:259–74, 2010
23. Donaldson T, Abbott, BR: Prediction in the individual case: an explanation and application of its use with the Static-99R in sexually violent predator risk assessments. J Am Coll Forensic Psychiatry 29:5–35, 2010
24. Vrieze SI, Grove WM: Multidimensional assessment of criminal recidivism: problems, pitfalls, and proposed solutions. Psychol Assess 22:382–95, 2010
25. Hanson RK, Bussiere MT: Predicting relapse: a meta-analysis of sexual offender recidivism studies. J Consult Clin Psychol 66:348–62, 1998

26. State of New Hampshire v. Thomas Hurley. Case No. 07-E-0236. 2009. Available at http://www.static99.org/pdfdocs/hurley-orderondaubert_nh.pdf. Accessed May 29, 2010

27. Hanson RK, Morton-Bourgon KE: The accuracy of recidivism risk assessments for sexual offenders: a meta-analysis of 118 prediction studies. Psychol Assess 21:1–21, 209

28. Meehl PE: Symposium on clinical and statistical prediction (with McArthur CC, Tiedeman DV). J Consult Psychol 3:163–73, 1956

29. Grove WM, Meehl PE: Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical-statistical controversy. Psychol Public Policy Law 2:293–323, 1996

30. Hanson RK, Morton-Bourgon KE: Predictors of sexual recidivism: an updated meta-analysis. Ottawa, ON, Canada: Department of the Solicitor General, 2004

31. Kingston DA, Seto MC, Firestone P, *et al*: Comparing indicators of sexual sadism as predictors of recidivism among adult male sexual offenders. J Consult Clin Psychol 78:574–84, 2010

32. Prentky RA, Janus E, Barbaree H, *et al*: Sexually violent predators in the courtroom: science on trial. Psychol Public Policy Law 12:357–93, 2006

33. Campbell TW: Assessing Sex Offenders: Problems and Pitfalls. Springfield, IL: Charles C Thomas, 2007

34. Mann RE, Hanson RK, Thornton D: Assessing risk for sexual recidivism: some proposals on the nature of psychologically meaningful risk factors. Sex Abuse J Res Treat 22:191–217, 2010

35. Beggs SM, Grace RC: Assessment of dynamic risk factors: an independent validation study of the violence risk scale: sexual offender version 2010. Sex Abuse J Res Treat 22:234–51, 2010

36. Hanson K, Harris AJR, Scott TL, *et al*: Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project. Available at http://www.publicsafety.gc.ca/res/cor/rep/_fl/crp2007-05-en.pdf. Accessed October 3, 2007

37. Eher R, Rettenberger M, Matthes A, *et al*: Dynamic risk assessment in sexual offenders using Stable-2000 and the Stable-2007: an investigation of predictive and incremental validity. Sex Abuse J Res Treat, in press

38. Eher R, Rettenberger M, Matthes A, *et al*: Stable dynamic risk factors in child sexual abusers: the incremental predictive power of narcissistic personality traits beyond the Static-99/Stable-2007 priority categories on sexual reoffense. Sex Offend Treat 5:1–12, 2010

39. Jackson R, Travia T, Schneider J: Annual survey of sex offender civil commitment programs. Sexual Offender Civil Commitment Network Research Committee. Available at http://www.soccpn.org/images/2008_20Annual_20Survey_20of_20Programs.pdf. Accessed October 25, 2009

40. Wollert R, Waggoner J: Bayesian computations protect sexually violent predator evaluations from the degrading effects of confirmatory bias and illusions of certainty: a reply to Doren and Levenson. Sex Offend Treat 1:1–23, 2009