

Running Head: Forensic Use of Static and Dynamic Risk Factors

Forensic Application of Static and Dynamic

Risk Factors: Is it the Right Time?

Brian R. Abbott, Ph.D., Independent Practice

111 N. Market Street. Suite 300

San Jose, CA, USA 95113

[brian@dr-abbott.net](mailto:brian@dr-abbott.net)

### **Nontechnical Summary**

Many places in the United States have laws that permit the involuntary civil confinement of individuals who have served their criminal sentences for having committed sexual crimes. Clinical forensic psychologists are the principal witnesses in these legal proceedings because they offer testimony as to whether individuals meet the requirements for involuntary civil commitment- one of which involves whether the person presents a certain likelihood of committing future sexual crimes. Psychologists have borrowed from research and theory about the assessment of sexual violence risk as applied clinically to address the legally defined likelihood for sexual reoffense under civil confinement law. The outcomes of clinically-based risk assessment procedures may not provide sufficiently reliable, relevant, or incisive information for a judge or jury to properly evaluate the legally defined likelihood to commit future sexual crimes. This article explores the challenges of applying two different forms of clinical sexual violence risk assessment as commonly applied by evaluators in involuntary civil confinement evaluations. Based on this analysis, recommendations for forensic practice and future research are offered.

### **Highlights**

- Assessment of the risk for sexual reoffense designed for clinical purposes may not provide sufficient evidence to address legal matters about the commission of future sexual crimes.
- I examine the validity and reliability of two methods of assessing sexual reoffense developed for clinical applications as applied to whether individuals present sufficient risk to commit future sexual crimes that warrants involuntary commitment as sexually violent predators or persons in the United States.
- The assessment of the likelihood for sexual reoffense under sexually violent predator person laws using the combination of sexual violence risk actuarial instruments and dynamic risk factors presents unique statistical and legal challenges some of which are surmountable and others that, in my view, are not.
- I make recommendations for forensic risk assessment and future research.

Data Availability: Not applicable

### **Abstract**

It is common, accepted clinical practice to conduct risk assessments of individuals who commit sexual offenses using the combination of sexual violence risk actuarial measures and dynamic risk factors. This assessment approach has utility when identifying treatment targets, assessing progress in sexual offender treatment, and forming risk management plans. Little research has examined this method in forensic contexts such as deciding whether individuals who suffer from mental disorders are likely to engage in sexually dangerous behavior as defined by sexually violent predator or persons (“SVP”) involuntary civil confinement laws in the USA. In particular, it is uncertain whether the combination of sexual violence risk actuarial measures and dynamic risk factors (DRF) produces sufficiently reliable, relevant, and probative evidence for the trier of fact to properly evaluate the SVP legally defined likelihood of sexual dangerousness. This article explores the efficacy of combining actuarial measures of sexual violence risk and dynamic risk factors as applied in SVP risk assessments based on some commonly observed forensic practices among evaluators. Based on the analysis, recommendations for forensic practice and future research are offered.

Keywords: Static-99R; STABLE-2007; sexual violent predator; sexual violence risk

## Introduction

It is generally recognized that dynamic risk factors (“DRF”) are indispensable when assessing the current potential for sexual reoffense among individuals undergoing sexual offender treatment or who are being supervised in the community for having committed sexual offenses (Association for the Treatment of Sexual Abusers, 2014; Olver et al., 2018; Hanson et al., 2007). It is recommended that DRF supplement static sexual violence risk actuarial instruments, such as the Static-99R, because DRF identify current, changeable psychological characteristics associated with sexual reoffense, which, in turn, can guide treatment planning and interventions and inform about methods to manage risk potential in the community (Mann et al., 2010; Association for the Treatment of Sexual Abusers, 2014; Phenix, Fernandez et al., 2016). Despite the widespread acceptance of applying DRF in clinical practice, little attention has been paid to the application of DRF in the forensic arena, particularly as it relates to the civil confinement of sexually violent predators or persons (“SVP”). The focus of this article is to critically examine the utility of combining static actuarial measures of sexual violence risk and DRF when determining whether individuals meet the likelihood to commit sexual reoffense proscribed by SVP statutes.

Laws in 21 states and federally permit the government to petition individuals for involuntary civil confinement as SVP after they have served their criminal sentences (Knighton et al., 2014). SVP statutes are premised on three underlying legal principles to justify involuntary civil confinement (Scurich & Krauss, 2014), including the existence of past qualifying criminal sexual conviction(s), the presence of a mental condition that causes serious difficulty controlling sexual behavior (“SVP mental disorder”), and the SVP mental disorder makes the person sexually dangerous. The only exception to this legal scheme is the federal Adam Walsh Act that

presumes the individual is sexually dangerous if he exhibits one of the qualifying sexual crimes and suffers from a current SVP mental disorder (Abbott, 2017). When civilly confined, the individual faces indefinite commitment unless he can later prove he no longer suffers from the SVP mental disorder or is no longer sexually dangerous (Scurich & Krauss, 2014).

The likelihood of sexual dangerousness in SVP statutes is defined by probabilistic language such as likely, substantially probable, and more likely than not (hereinafter referred to as “likely”). The probabilities of sexual reoffense predicted by sexual violence risk actuarial measures are recognized as providing sufficiently relevant and probative evidence for the fact finder to evaluate whether individuals being petitioned as SVP meet the likely threshold (Abbott, 2017; Duwe & Kim, 2016; Helmus et al., 2012; Janus & Prentky, 2003; Prentky et al., 2006; Woodworth & Kadane, 2004). It is common practice for forensic examiners for the state to employ sexual violence risk actuarial measures (Jackson & Hess, 2007; Schneider et al., 2014) in combination with dynamic risk factors (Sreenivasan et al., 2010) to assess whether individuals meet the likely threshold. This method of risk assessment I refer to as the adjusted actuarial approach (“AAA”) and its use has been advocated in SVP risk assessments elsewhere (Sreenivasan et al., 2010; Abbott, 2011). I am not aware of surveys or studies that have examined the application of the AAA in SVP assessments, but in my review of hundreds of SVP forensic reports authored by evaluators from nine states, a common practice emerges.

Evaluators implement the AAA by combining one or more sexual violence risk actuarial measures with DRF. One of two procedures is typically employed to assess DRF. One method involves the completion of standardized measures such as the STABLE-2007 (Fernandez et al., 2014) or VRS-SO DRF (Olver et al., 2018), hereinafter referred to as the standardized AAA. In the second process, the evaluator selects and weighs DRF that are obtained from meta-analytic

research (e.g., Mann et al., 2010; Hanson & Morton-Bourgon, 2004; Hanson & Bussière, 1998), which I will refer to as the unstructured AAA. In rare instances, evaluators may employ both standardized and unstructured approaches to assess DRF. The theory of the standardized or unstructured AAA as applied in SVP risk assessments assumes the probabilities for sexual reoffense predicted by sexual violence risk actuarial measures are insufficient to satisfy the likely threshold. It is further believed that the combination of the results from sexual violence risk instruments and the measures of DRF produce probabilities of sexual reoffense sufficient to substantiate that evaluatees meet the likely threshold (Abbott, 2011). In practice, evaluators report the known probability for sexual reoffense estimated by the sexual violence risk measure over a specific follow up period and then state the DRF found present increases the evaluatee's likelihood to reoffend sexually beyond the known predicted probability, although the extent of the increase is not stated and the revised probability estimate is not quantifiable. Some standardized AAA procedures permit the evaluator to report probabilities of sexual reoffense based on the outcome of the combined the static and dynamic instruments.

In practice, I commonly see the standardized AAA consisting of Static-99R and STABLE-2007 (Brankley et al., 2017) or the Static-99R and VRS-S0 DRF (Olver et al., 2018). The research supporting these risk assessment procedures do not report whether the probabilities of sexual reoffense from the joint measures are significantly greater than the score-wise rates reported by Static-99R actuarial tables (Phenix et al., 2016). This difference can be discerned by post-hoc analysis that I will discuss in the section on conclusions and recommendations for forensic practice. Nonetheless, I find that evaluators commonly eschew reporting the objective data in favor of rendering a qualitative conclusion that the DRF found present for the evaluatee increases the likelihood for him to commit future sexual offending at some unquantifiable rate

that is greater than the score-wise rate from the sexual violence risk actuarial instrument.

Empirical support lacks to support this inference.

A recent meta-analytic study by van den Berg et al. (2018) examined the incremental predictive validity of DRF instruments over static measures of sexual violence risk. They examined 13 unique samples with an aggregate of 3,747 individuals. Random effect meta-analysis revealed that the DRF instruments produced statistically significant incremental validity over static measures of sexual violence risk (HR = 1.09; 1.06, 1.12) with moderate chance variability in effect sizes across the 13 studies (van den Berg et al., 2018). The small effect size may have resulted from redundancy between the static and dynamic risk factors (van den Berg et al., 2018). Ward and Beech (2015) argue that static and dynamic risk factors may be measuring the same underlying risk propensities (e.g., sexual deviance and antisociality), but do so in different ways, and this would likely explain the small increase in variance associated with the measures of dynamic risk. Since dynamic risk factors produce a modest increase in the amount of variance accounted by the sexual violence risk actuarial instruments, I think it is reasonable to infer there would likely be a reciprocal effect on the increase in the likelihood of sexual reoffense over that predicted by actuarial measures alone. Unfortunately, in my view, the research is lacking to test this hypothesis.

What follows is an exploration of whether the clinical practice of considering an actuarial measure of sexual violence risk and DRF to identify treatment targets or to manage risk in the community directly translates to a forensic evaluator's task to produce sufficiently reliable, relevant and probative evidence, in the form of probabilities for sexual reoffense, so the fact finder can properly evaluate whether individuals being prosecuted as SVP meet the likely threshold. I am not questioning the established predictive accuracy of DRF or their use for

clinical purposes. Rather, I address whether the standardized and unstructured AAA produce quantifiable probabilities of sexual reoffense that are meaningfully greater than that predicted by the actuarial instrument alone as purported by evaluators who use these procedures. The issue is whether outcomes of the standardized and unstructured AAA generate sufficiently reliable, relevant, and probative evidence that permits the trier of fact to appropriately evaluate whether an individual meets the SVP likely threshold. Consistently with this goal, I will critically examine the assumption that the application of sexual violence risk actuarial measures in combination with standardized or unstructured measures of DRF actually produce quantifiable probabilities of sexual reoffense that are greater than those predicted by the actuarial measure alone. The concepts I present and discuss below would apply to the combination of any measure of DRF and sexual violence risk actuarial instrument, but for the sake of clarity when illustrating the issues, I will limit the presentation to the common forensic practice among evaluators who use of the Static-99R, which happens to be the most commonly actuarial measure is SVP risk assessments (Jackson et al., 2008), other psychological risk factors from Mann et al. (2010), and the STABLE-2007 (Fernandez et al., 2014).

### **Unstructured Application of DRF**

Since 1998, two meta-analytic studies have been published (Hanson & Morton-Bourgon, 2004; Hanson & Bussière, 1998) that examined static and dynamic risk factors associated with the occurrence of sexual reoffense. Mann et al. (2010) published a meta-analysis that specifically examined other psychological risk factors that were comprised of dynamic risk factor and long-term vulnerabilities for risk. All three studies employed an effect size statistic (Cohen's *d* or correlation) to measure the extent of the relationship between a risk variable and sexual violence risk (i.e., univariate relationship). The studies revealed that effect sizes for most of the dynamic

risk factors were small (i.e., Cohen's  $d < .50$ ). None of the meta-analyses examined multivariate relationships between two or more DRF and sexual reoffense or the combination of DRF and sexual violence risk actuarial instruments.

Evaluators employing the unstructured AAA substitute this procedure in place of or in addition to the application of a standardized DRF measure such as the STABLE-2007 or VRS-S0. The evaluator selects a predetermined number of DRF that may be as few as three and as great as a dozen. The hallmark of this approach is its variation within and across evaluators. An evaluator may tailor the selection of DRF to assess clients based on unique circumstances of the cases, while some evaluators rely upon an identical list of DRF for all risk assessments. The unstructured AAA assumes that the presence of the first DRF increases the likelihood of sexual reoffense over that predicted by the sexual violence risk actuarial measure by some unspecified and actually unknown magnitude. It is further presumed that the sexual recidivism rate successively increases by an unknown magnitude for each other DRF found present for the evaluatee. Table 1 illustrates a common list of dynamic risk factors considered by some evaluators that was obtained from Mann et al. (2010; Table 2, supported and promising variables).

**Table 1: List of Dynamic Risk Factors with Associated Statistical Outcomes**

Risk Factor	Present	Mean d	AUC	1-AUC
Sexual preoccupation	Yes	.39	.61	.39
Sexual preference for children (PPG)	No	.32	.58	.42
Sexualized violence	Yes	.18	.55	.45
Offense-supportive attitudes	Yes	.22	.56	.44
General self-regulation problems	Yes	.37	.60	.40
Poor cognitive problem solving	Yes	.22	.56	.44

Noncompliance with supervision	Yes	.62	.67	.33
Grievance/hostility	No	.20	.55	.45
Negative social influences	Yes	.26	.57	.43
Machiavellianism	Yes	1.4	.84	.16
Callousness/lack of concern for others	Yes	.29	.58	.42

Columns 1 and 2 in Table 1 reflect the DRF a hypothetical evaluator selects to apply to the evaluatee and the outcome of the assessment for each factor (i.e., present or absent). In this example, the evaluator determined that 9 out of the 11 DRF considered were applicable to the evaluatee. Column 3 reports the associated mean Cohen's *d* for each DRF as reported by Mann et al. (2010). According to Cohen (1988), the effect size *d* can be interpreted as follows: small ( $d = 0.20$ ), medium ( $d = 0.50$ ), and large ( $d = 0.80$ ). Consistent with the results from Mann et al. (2010), the majority of the DRF in Table 1 achieved small effect sizes, with one reaching a moderate effect size (.62) and another attaining a large effect size (1.4). The area under the curve (AUC) values reported in column 4 were obtained by transforming the Cohen's *d* into the corresponding AUC using a conversion table published by Salgado (2018). The AUC represents the correct classification of recidivists who exhibited the DRF. The last column reports the values corresponding to 1 minus the AUC value or the false positive rate (Streiner & Cairney, 2007), which in this situation represents the proportion of nonrecidivists who exhibited the DRF but were misclassified as recidivists.

The unstructured AAA presumes that each risk factor present not only contributes unique variance, but it also increases the probability of sexual reoffense over other DRF considered and the Static-99R. This assumption is simply unsupported by the meta-analytic studies evaluators rely upon to justify the procedure (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2004;

Mann et al., 2010). The meta-analytic studies examined the univariate relationships only (i.e., between a single DRF and sexual violence risk). Therefore, the results from the meta-analyses cannot be relied upon to assess the unique or incremental contribution of two or more DRF as it relates to sexual violence risk (Mann et al., 2010). Similarly, the three meta-analyses did not examine the extent to which one or more DRF account for unique or incremental variance beyond that covered by a sexual violence risk actuarial measure.

The unstructured AAA further posits that the selected DRF that are identified as present with the evaluatee lead to only one outcome, that is, an increased potential to reoffend sexually. For this conclusion to be true, one has to accept the notion that each DRF perfectly discriminates between recidivists and nonrecidivists. This premise is unsupported as illustrated in Table 1. The AUC values for each DRF indicate the extent to which randomly selected recidivists are discriminated from the randomly selected nonrecidivists and none reach perfect discrimination. This means that a certain proportion of individuals who presented with a specific DRF did not reoffend sexually and subtracting the AUC values from 1.0 reveals this information. With the exception of Machiavellianism, there is about a 40% to 45% probability that a given nonrecidivist will exhibit the DRF, but they will then be misclassified as sexual recidivist when using the unstructured AAA. The meaningful potential for misclassifying nonrecidivists as recidivists, the possible redundancy across the selected DRF, and the extent of shared variance between DRF with sexual violence risk actuarial instruments cannot be accounted for by forensic practitioners' professional judgment. For this reason, the unstructured AAA introduces an unknown magnitude of error in decision making. To date, there have been no scientific studies to test the predictive validity of the unstructured AAA, but two studies have examined the

predictive validity of the standardized AAA, and the results illuminate the limitations and challenges of the unstructured AAA.

Vrieze and Grove (2010) attempted, to no avail, to combine various sexual violence risk actuarial measures with standardized professional judgment instruments, one of which contained dynamic risk factors. The researcher encountered six pitfalls that would need to be overcome before producing accurate results when combining different risk assessment instruments. The same pitfalls would occur when using the unstructured AAA. Vrieze and Grove (2010) argue that it is unlikely that a risk assessment approach like the unstructured AAA would incrementally increase rates of discrimination accuracy for sexual violence risk actuarial instruments because, *“One expects, given clinician (i.e., human) fallibility in determining base rates, scoring instruments, applying cutting scores, combining the results from diverse tests, and making clinical/professional judgments during the entire process, that long-term clinical field accuracy will fall short of AUCs reported in the literature [for actuarial measures].”*

Mokros et al. (2010) utilized the multivariate Bayesian classification statistical method to test the standardized AAA. While they addressed violent reoffending using standardized professional judgement measures, the study methodology would apply equally when evaluating the predictive accuracy of the unstructured AAA. The multivariate Bayesian classification procedure controls for the base rate of reoffense, the extent to which the risk factors discriminate between recidivists and nonrecidivists, and the redundancy of risk factors. The researchers tested 255 possible combination of risk factors and discovered that two (age and factor 2 total score from the PCL-R) provided the greatest selection accuracy as measured by the AUC. The result of this study suggests that many of the risk factors contained in Table 1 would not improve predictive accuracy or increase the probability of sexual reoffense predicted by the sexual

violence risk actuarial measure. In fact, because the base rate of sexual violence risk is generally lower than the violent reoffense rate that Mokros et al. (2010) used as the recidivism criterion, it is reasonable to conclude that it would be less likely for DRF to accurately separate sexual recidivists from nonrecidivists.

Some evaluators conclude that the presence of each DRF for an individual is associated with a specific increase in the probability of sexual reoffense over those individuals who do not exhibit the DRF. This approach appears to have its roots in the Hanson and Bussière (1998) meta-analysis where the researchers stated the correlation coefficient for a risk factor could be interpreted as the difference in the probability for sexual reoffense between the group with the risk factor and the group without it centered around the base rate. For example, if the base rate of sexual violence risk was 25% in a group of persons convicted for sexual offenses and the correlation with noncompliance with supervision from Mann et al. (2010) is .30, then the sexual violence risk rate for the group with the risk factor would be 40% and 10% for the group lacking the risk factor. It becomes intuitively apparent that this effect, which is referred to as the binomial effect size display (“BESD;” Lipsey & Wilson, 2001; Randolph & Edmondson, 2005), yields implausible results in low base rate conditions. Using the same example cited previously, but adjusting the base rate to 10%, produces a 25% sexual reoffense rate for those individuals who lack cooperation with supervision and -5% in those without it. This impossible result occurs because the differences in sexual violence risk rates derived from BESD assumes a 50% base rate (Lipsey & Wilson, 2001; Randolph & Edmondson, 2005). Unless the individual being assessed is a member of a group of individuals who commit sexual offenses and who sexually reoffend at rate that does not dramatically depart from 50%, the application of the BESD would be improper. Even if circumstances existed to consider the BESD, the results only apply to the

effect for a single variable. One would be hard pressed to find an empirically defensible way to systematically integrate the BESD results across multiple DRF or with the probability estimate generated by sexual violence risk actuarial measure.

The promising risk factors listed in Table 1 list two variables with limited scientific support, including Machiavellianism and callousness/lack of concern for others. The two DRF have limited support in that the former was identified in a single study of 99 child molesters from a prison treatment program in the United Kingdom (Thornton, 2003) and the latter is supported by two studies (Hanson et al, 2007; Knight & Thornton, 2007). The reliability of the effect size from studies supporting the two promising DRF remains uncertain pending as replication studies. Additional research will inform as to whether the observed effects sizes remain stable across samples, decrease, or increase. Moreover, it is reasonable to infer that the results from a single study of 99 men treated for sexual offending in prison in the United Kingdom more than 25 years ago are of questionable validity when applied to this population in present day.

The Machiavellianism DRF illustrates an issue about the criteria used to identify the DRF and whether the same procedure is used in present day risk assessments. Mann et al. (2010) describe persons who exhibit Machiavellianism as viewing others as weak, cowardly, and selfish, and, therefore, it is appropriate to take advantage of others. Mann et al. (2010), however, do not provide direction as to a valid and reliable procedure to assess these personality characteristics nor would this be expected based on methods of meta-analytic research. Inspection of the source documentation (Thornton, 2003) reveals the researcher modified an instrument known as the MACH-IV (Christie & Geis, 1970) to assess the presence of Machiavellianism. The nature of the alteration to MACH-IV was not reported. Evaluators who use rating criteria other than the modified MACH-IV may lack a valid basis by which to identify

the extent to which individuals exhibit this DRF. Indeed, it is uncertain what evaluators are actually measuring when assessing Machiavellianism by methods other than the modified MACH-IV.

The discussion above raises a general issue about the reliability and validity of the unstructured AAA approach. Brief descriptions of DRF contained in meta-analytic studies do not provide for standardized rating criteria by which to assess individuals that are the hallmark of valid and reliable measurement (American Educational Research Association et al., 2014). Research has demonstrated how the effect of partisan allegiance in high stakes legal matters, such as SVP evaluations, may degrade the reliability of outcomes when administering well-researched measures, such as the Static-99R or PCL-R (Murrie et al., 2009; Boccaccini et al., 2009), despite these instruments showing good field reliability in other forensic contexts (Boccaccini et al., 2012; Olver et al., 2020). It is reasonable to infer that rating DRF without established valid rating criteria in high stake SVP forensic evaluations will in all likelihood produce results with unacceptable levels of error (i.e., false positive and false negative outcomes).

Support for the unstructured AAA is grounded in studies establishing the predictive accuracy of DRF based on discrimination statistics (e.g., AUC or correlation coefficient). Discrimination statistics inform as to how well an instrument separates those who reoffend sexually from those who do not (Cook, 2007), but this is the improper analysis to assess the accuracy of the unstructured AAA in SVP risk assessments. The task of the evaluator is to decide whether the unstructured AAA as applied to the evaluatee generates a likelihood of committing future criminal sexual acts that meets or exceeds the SVP likely threshold. The accuracy of this

fit is a matter of calibration rather than discrimination. I am not aware of any calibration studies having been conducted to test the unstructured AAA in SVP risk assessments.

### **Standardized AAA**

The standardized AAA involves the administration of a sexual violence risk actuarial measure and a standardized instrument that has been developed and validated for measuring dynamic risk factors such as the STABLE-2007 (Hanson et al., 2007) or VRS-SO dynamic risk measure (Olver et al., 2018). I will address the standardized AAA by examining the use of the Static-99R and the STABLE-2007. As presented below, research regarding the two instruments has involved the Static-99 and Static-99R, but for ease of presentation I will use the designation of Static-99R. I will cover three major areas in this section, as it relates to persons who are petitioned for civil commitment as SVP, including the proper administration of the STABLE-2007, the predictive validity of STABLE-2007, and interpretation of the combined results from the Static-99R and STABLE-2007.

While the STABLE-2007 was originally devised on a population of individuals under community supervision for sexual offenses (Hanson et al., 2007), eight studies have examined the administration of the STABLE-2007 when the persons were in custody for sexual offenses and they were later released into the community. This group fits individuals who are petitioned for civil confinement as SVP because they are held in custody pending the legal determination whether to involuntarily confine them. If not civilly committed, they are released into the community. Table 2 presents the results from the eight studies that examined the incremental predictive validity of the STABLE-2007 over the Static-99R. For each study, Table 2 describes its geographic location and the total sample size, as well as whether the STABLE-2007 was significantly associated with sexual violence risk, and whether the STABLE-2007 achieved

incremental predictive validity over the Static-99R. The values for the associated measures of predictive validity are also specified, including the AUC statistic and Beta from either Logistic regression or Cox regression survival analysis. It is noteworthy that the results reported in Table 2 represent discrimination accuracy. The ability of the instruments to sort recidivists from nonrecidivists is not the relevant metric for psychologists or the triers of fact to weigh whether an individual's likelihood of sexual reoffense meets the legally defined threshold of "likely." The central issue when relying upon the standardized AAA for SVP risk assessments is whether calibration evidence produces likelihoods of sexual reoffense from the combined measures that are significantly greater than those rates predicted by the sexual violence risk instrument alone.

**Table 2: Summary of Results from Studies Examining the STABLE-2007 Alone and in Combination with the Static-99R**

Study	Country	N	Sexual Recidivism	
			Did STABLE-2007 Predict? (Measure)	Did STABLE-2007 and Static-99R Predict? (Measure)
1. Saum (2007)	USA	175	Yes (AUC = .68)	No ( <i>B</i> not reported)
2. Eher et al. (2012)	Austria	264	Yes (AUC = .71)	No ( <i>B</i> = .11)
3. Looman & Abracen (2012)	Canada	168	Not reported	Yes ( <i>B</i> = .095)
4. Eher et al. (2013)	Germany/ Austria	370	Yes (AUC = .71)	Yes ( <i>B</i> = .18)
5. Eher et al. (2015)	Austria	189	No (AUC = .60)	No (AUC = .64)
6. Looman & Goldstein (2015)	Canada	442	Yes (Beta = .23)	No ( <i>B</i> = .09)
7. Sowden & Olver (2017)	Canada	180	No (AUC = .56)	No, pretreatment ( <i>B</i> = .06) Yes, posttreatment ( <i>B</i> = .08)
8. Etzler et al. (2018)	Austria	638 <sup>a</sup>	Yes (AUC = .64)	No by total score ( <i>B</i> = .052) Yes by risk category ( <i>B</i> = .504)

<sup>a</sup> This sample contains the 264 individuals from Eher et al. (2012).

As seen in Table 2, seven of the eight studies tested whether the STABLE-2007 predicted sexual violence risk via the AUC or by Cox Regression. All but two of the studies discovered that the STABLE-2007 achieved statistically significant predictive accuracy for sexual reoffense. The ability of the STABLE-2007 to add to the prediction of sexual violence risk beyond the Static-99R revealed mixed results. Of the seven studies that considered the STABLE-2007 total score, not moderated by treatment status, five discovered the STABLE-2007 did not contribute additional unique variance. When combining STABLE-2007 total scores into risk categories of low (0-3), moderate (4-11), and high ( $\geq 12$ ), Etzler et al. (2018) found that the instrument produced incremental predictive validity over the Static-99R alone. Results from Sowden and Olver (2017) revealed that the time at which the STABLE-2007 was administered (post treatment) influenced whether it achieved incremental predictive validity. It is uncertain what contributes to the tendency for the STABLE-2007 failing to achieve significant incremental predictive validity over the Static-99R alone, however, recent meta-analytic results by Brankley et al. (2019) appears to shed light on this issue and this will be addressed in the discussion section.

The studies reported in Table 2 completed the Static-99R and Stable-2007 according to standardized administration and scoring procedures (Fernandez et al., 2014; Phenix, Fernandez et al., 2016). Contrary to these expected standards of test administration and interpretation, I have observed evaluators complete the STABLE-2007 in nonstandardized ways among which may include ignoring the item rating instructions, altering item scores (e.g., present/absent, aggravate/not aggravate, or presence of risk factor does not aggravate risk), failing to compute the STABLE-2007 total score, or considering only the effect of dynamic risk items deemed present. Such idiosyncratic procedures deviate substantively from how the STABLE-2007 was

designed, validated, and replicated, which invalidates the results and prevents drawing conclusions about incremental predictive validity of the Stable-2007 over the Static-99R.

In summary, the majority of the studies reported in Table 2 found the STABLE-2007 to achieve moderate discrimination accuracy among individuals in custody who were convicted of sexual offenses. Less support has been found for the proposition that the STABLE-2007 accounts for more variance than the Static-99R alone among in-custody samples. The studies from Table 2 that established statistically significant discrimination accuracy did not examine calibration of the Static-99R along with the combined measures. I find it common practice for some evaluators to assume wrongly that statistically significant discrimination accuracy for the combination of the Static-99R and STABLE-2007 ipso facto results in a higher probability of sexual reoffense than predicted by the Static-99R alone. Setting aside that such a conclusion conflates discrimination accuracy with calibration, empirical evidence lacks to determine to what extent, if any, the observed probability of sexual violence risk for the combined instruments is greater than that predicted by the Static-99R alone. To examine this issue, I obtained two data sets where the STABLE-2007 achieved statistically significant discrimination accuracy over the Static-99R. I conducted a separate calibration analysis to discover whether the sexual reoffense rates for the combined instruments were greater than the observed sexual recidivism rate from the Static-99R alone. The following describes the methodology and results of this exploratory calibration analysis.

*Participants:* One dataset consists of 566 Canadian men convicted for sexual offenses who were part of the dynamic supervision project (Helmus & Hanson, 2013). The sample is comprised of individuals who were under community supervision related to being convicted of sexual offenses at the time the STABLE-2007 was completed. Detailed information about the

study methodology and sample can be found in Hanson et al. (2007). In brief, the start dates in the community ranged between January 18, 2001 and October 19, 2006, with a median follow up time of 41 months ( $M = 40.9$ ,  $SD = 13.3$ , range = 1 - 65 months). Sexual recidivism was defined as all crimes of sexual motivation regardless of whether the charged offense was explicitly sexual, including official charges, self-reported reoffense, or breaches of supervision resulting in parole revocation or conviction for violation of conditional release. The base rate of sexual reoffense over a fixed five-year period was 11.7%. A total of 513 individuals fit the requirement for a fixed five-year follow-up period. Predictive validity was tested using the AUC. Incremental predictive validity was calculated using Cox regression. The STABLE-2007 was significantly predictive of sexual recidivism ( $AUC = .67$ ,  $95\%CI = .60, .74$ ), as well as incrementally predicting this outcome beyond the Static-99R alone ( $HR = 1.075$ ;  $p = .003$ ).

The second sample I analyzed came from a study conducted by Looman and Goldstein (2015). Looman provided the data, as well as additional information about the study methodology and results at my request, which is presented below. The sample consisted of two groups of individuals who were treated at the Ontario Region of the Correctional Service of Canada with an aggregate total of 442 sexual offenders. The first group of 376 subjects had the Static-99R and the STABLE-2000/2007 completed as part of specialized in-custody sexual offender assessment that was completed within three to five months of their reception to the Correctional Service of Canada. Of these men, 247 completed a sexual offense treatment program during their sentence, 43 refused treatment, 22 were discharged from treatment prior to completion (typically for failure to comply with program rules), and no evidence of being offered treatment prior to release was discovered in records of 24 men. Data concerning the treatment status for the remaining 40 individuals was not available. Information used to score the

instruments included police reports and court documents related to their trial/sentencing and when available presentence reports, psychological/psychiatric assessments completed prior to sentencing and any documents available for those who had previous sentences.

The second group consisted of 66 men who were assessed as part of the pretreatment assessment for the Regional Treatment Centre-Sex Offender Treatment Program (RTC-SOTP; Abracen & Looman, 2015). This group consisted of men who entered the correctional system prior to the use of the STABLE on intake; however, the STABLE 2000/2007 was scored as part of the pretreatment assessment for the RTC-SOTP. The treatment status of this group was as follows: 3 were assessed only; 45 completed treatment; 14 were discharged from treatment; and 4 withdrew from treatment. Information used to score the dynamic risk measures included the previously mentioned sources, as well as any information which became available while serving their sentence before entering treatment among which may have included reports from other programs and reports regarding behavior during institutional employment.

A total of 350 men were followed for the entire fixed five-year period. Recidivism data were collected on the subjects from official criminal records maintained by the Royal Canadian Mounted Police (RCMP). The official Fingerprint Service (FPS) sheets for each case was obtained electronically and new convictions were coded according to the Cormier-Lang system (Harris, Rice, Quinsey & Cormier, 2015). New sexual offenses were those offenses clearly of a sexual nature according to the recorded conviction (e.g., sexual assault, gross indecency, invitation to sexual touching). Outcome data was collected during the summer of 2014. The average follow-up time was 6.1 years ( $SD = 2.9$ ; range = 6 days to 12.9 years). The fixed five-year sexual reoffense base rate was 4.3%. Cox regression was used to examine predictive validity and IPV. The STABLE-2007 was significantly predictive of sexual recidivism via Cox

regression ( $\text{Exp}(B) = 1.17$ ;  $p < .001$ ), and also incrementally predicted this outcome over the Static-99R alone ( $\text{Exp}(B) = 1.11$ ;  $p = .009$ ).

*Method:* Based on the statistically significant discrimination accuracy of many of the studies examining the STABLE-2007 and Static-99R with individual in-custody for sexual offending, I explored whether the combined measures produced observed sexual reoffense rates that were significantly greater than that predicted by the Static-99R alone according to the following procedure. The Static-99R total scores were grouped into risk bins, based on the procedure described by Fernandez et al. (2014), with the corresponding Static-99R total scores for each category listed in parenthesis: low (-3 to 1), moderate (2, 3), Moderate-High (4, 5), and High ( $\geq 6$ ). The bin-wise sample sizes and numbers of recidivists were used to calculate the observed bin-wise probability of sexual reoffense. The sexual reoffense rates are reported for a fixed five-year follow-up. The STABLE-2007 scores were grouped into risk bins, as specified by Fernandez et al. (2014), with the range of STABLE-2007 total scores following in parentheses, including low (0 – 3), moderate (4 – 11), and high ( $\geq 12$ ). Based on instructions from Fernandez et al. (2014), the combination of the Static-99R and STABLE-2007 risk bins produce corresponding Static/STABLE priority categories of risk, including low, moderate-low, moderate-high, high, and very high. Each Static-99R and STABLE-2007 risk bin combination formed between two and three Static/STABLE priority categories (see Tables 3 and 4 for the designated priority categories associated the combination of Static-99R and STABLE-2007 risk bins). The five-year observed sexual reoffense rate was computed for each Static/STABLE priority category.

The sexual reoffense rate for each Static-99R risk bin was compared to the sexual recidivism rates from the corresponding Static/STABLE priority categories. Consistent with the

rationale for standardized AAA, it is hypothesized that the sexual reoffense rate for the Static/STABLE priority category would be greater than the observed rate of sexual reoffense for the corresponding Static-99R risk bin at a statistically significant level. Testing these differences was accomplished by using a z-test for comparing two proportions using a SPSS v.24 macro. The result of the macro is essentially the same as the 2 x 2 chi-square, with the advantage of the latter being the inclusion of the phi correlation effect size. Thus, for all analyses the chi-square test was used to compare the proportions using a level of significance of  $\alpha = .05$ . Though judging the magnitude of effect size is, in part, context dependent, for this study a small, medium, and large effect size for the phi correlation correspond to 0.10, 0.24, and 0.50, respectively (Harris & Rice, 2005).

*Results:* Table 3 reports the results of the comparison from the dynamic supervision project (“DSP”) for all raters. I had data for a subsample of conscientious raters (all Canadian raters), but the results were the same as for all raters. Interested readers can request the conscientious raters’ results from this author. The Chi-Square analysis revealed that the observed sexual recidivism rates for Static/STABLE priority categories were not significantly different than the corresponding Static-99R risk bins. All effect size correlations indicate that the magnitude of the differences in sexual recidivism rates between the Static/STABLE priority categories and the corresponding Static-99R risk bins were small. The results are contrary to the study hypothesis that the recidivism rates for the combined measures would be significantly greater than the Static-99R alone.

**Table 3: Static-99R and STABLE 2007 5-Year Follow Up DSP: All Raters**

Static-99R Bin & Associated Total Scores	R <sup>+</sup> /N	Recidivism Rate	STABLE-2007 Score Groups	Priority Category	R <sup>+</sup> /N	Recidivism Rate*	Effect Size Phi
Low Score ≤ 1	12/182	6.6%	Low & moderate	Low	11/165	6.6%	.001
			High	Mod-Low	1/17	5.9%	-.008
Moderate-low 2, 3	10/171	5.8%	Low	Low	0/38	0.0%	-
			Moderate	Mod-low	7/99	7.1%	.024
			High	Mod-high	3/34	8.8%	.045
Moderate-High 4, 5	17/102	16.7%	Low	Mod-low	1/5	20.0%	.019
			Moderate	Mod-high	8/60	13.3%	.045
			High	High	8/37	21.6%	.057
High ≥ 6	21/58	36.2%	Low & moderate	High	8/25	32.0%	-.04
			High	Very High	13/33	36.4%	.032

\* All Chi-Square analyses were nonsignificant

The results reported in Table 4 are consistent with what was revealed in Table 3. The Chi-Square analysis found that the observed sexual recidivism rates for the Static/STABLE priority categories were not significantly different than the corresponding Static-99R risk bins. The magnitude of the differences, as measured by the effect size statistic, were negligible for all but one Static/STABLE priority category. The 20.6% sexual recidivism rate at the very high Static/ STABLE priority category reflects a small magnitude of difference compared to the corresponding 11.7% sexual recidivism rate for the Static-99R high risk bin. Nonetheless, the overall results do not support the study hypothesis that the recidivism rates for the combined measures would be significantly greater than the Static-99R alone.

**Table 4: 5-Year Sex Recidivism Rates from Looman and Goldstein (2015) Data**

Static-99R Bin & Associated Total Scores	R <sup>+</sup> /N	Recidivism Rate	STABLE-2007 Score Groups	Priority Category	R <sup>+</sup> /N	Recidivism Rate*	Effect Size Phi
Low Score ≤ 1	4/165	2.4%	Low & moderate	Low	4/163	2.5%	.001
			High	Mod-Low	0/2	0.0%	-
Moderate-low 2, 3	3/65	4.6%	Low	Low	0/10	0.0%	-
			Moderate	Mod-low	2/43	4.7%	.001
			High	Mod-high	1/12	8.3%	.061
Moderate-High 4, 5	1/60	1.7%	Low	Mod-low	0/36	0.0%	-.079
			Moderate	Mod-high	1/23	4.3%	.078
			High	High	0/1	0.0%	-.017
High ≥ 6	7/60	11.7%	Low & moderate	High	0/26	0.0%	-.196
			High	Very High	7/34	20.6%	.12

\* All Chi-Square analyses were nonsignificant

There are limitations resulting from the study methodology that may have affected the results. The small cell sizes and low base rates likely created instability in the observed sexual recidivism rates (Hanson et al., 2017; Olver et al., 2018). The Phi effect size statistic may be attenuated by both the loss of variance and low base rates, which likely effects the detection of the magnitude of difference. Tables 3 and 4 illustrate that simple priority tables are too problematic to use because of their reliance on observed sexual recidivism rates with some small cell frequencies. This analysis indicates that higher powered statistical analysis is indicated to address the calibration of combined static and dynamic risk measures, which is discussed in the recommendations for future research.

## Discussion

It is standard clinical practice to apply a combination of sexual violence risk actuarial measures and DRF, in unstructured ways or via standardized instruments, when assessing sexual recidivism risk of individuals who commit sexual offenses (ATSA, 2014; Olver et al., 2018; van

den Berg et al., 2018). The analyses I presented herein do not contest the use of DRF in this fashion. Rather, I examined whether the clinical application of DRF meets the demands of forensic risk assessment of individuals who are petitioned for involuntary civil confinement under SVP statutes. The utility of the unstructured and standardized AAA in SVP risk assessment is premised on a body of research demonstrating that individual DRF or standardized measures of dynamic risk produce moderate discrimination accuracy, as well as DRF contributing unique variance over sexual violence risk actuarial measures alone. The same rationale forms the basis of some evaluators' conclusions that DRF found present in an evaluatee increases the probability of sexual reoffense over the rate determined by the sexual violence risk actuarial measure alone, but such reasoning is in error because it conflates discrimination with calibration. Calibration studies have not been conducted to test the validity of this purported outcome.

The unstructured AAA essentially lacks scientific support and it amounts to using professional judgment to adjust the results generated by the sexual violence risk actuarial measure, which is known to reduce predictive accuracy (Duwe & Roque, 2018; Hanson & Morton-Bourgon, 2009; Storey et al., 2012; Wormith et al., 2012). Calibration studies lack to show the unstructured AAA produces sexual recidivism estimates that are greater than the probabilities for reoffense predicted by sexual violence risk actuarial measures alone. Therefore, in my opinion, evaluators rely upon speculation when they testify or conclude in reports that the consideration of selected DRF found present in the evaluatee produces a likelihood of sexual reoffense that is greater, by some unstated magnitude, than the rate predicted by the sexual violence risk actuarial measure alone. When faced with such speculative risk assessment testimony, an SVP mock juror study (Krauss & Scurich, 2013) suggests that the trier of fact may

be unduly influenced to reach a verdict for civil confinement. This raises the question as to whether such unsupported and unelaborated testimony should be admissible at trial (Scurich & Krauss, 2013), as well as the propriety of psychologists reporting or testifying about assessment results whose validity and reliability have not been established with evaluatees undergoing SVP commitment evaluations (American Psychological Association, 2013).

The results from Mokros et al. (2010) multivariate Bayesian classification analysis appears to support the theory of Occam's Razor, where the least number of risk factors provide optimal discrimination of recidivists from nonrecidivists. The idea that the accuracy of prediction is inversely related to the number of predictor variables has been previously noted (Seto, 2005) and rebuts the rationale in support of the application of the unstructured AAA in SVP risk assessments, which is premised on improving predictive accuracy by accounting for as many possible sources of sexual violence risk. I am not advocating that the unstructured AAA approach be abandoned altogether. It has clinical utility such as identifying targets of intervention for individuals undergoing sexual offender treatment and for evaluating progress in treatment. DRF presented by an individual who is released in the community can become the basis for guiding supervision practices to reduce sexual reoffense potential. These applications of the unstructured AAA may also be applicable for individuals judicially committed as SVP when they participate in sexual offender treatment or when they are later conditionally released into the community. Applications of the unstructured AAA for clinical and risk management of SVP; however, do not justify using it when assessing the likely threshold in legal proceedings to determine whether individuals meet the likely threshold that justifies involuntary civil detention.

Studies that examine individuals who are in custody for sexual offending reveal a trend toward the STABLE-2007 having moderate accuracy in separating recidivists from

nonrecidivists, however, the combination of the STABLE-2007 and Static-99R was less consistent in discrimination accuracy. No studies have been published that have examined whether the standardized AAA involving the STABLE-2007 and Static-99R produce sexual recidivism estimates that are significantly greater than that predicted by the Static-99R alone. The results from the exploratory study suggest the observed sexual recidivism rates based on the STABLE-2007 priority levels do not differ substantially from the observed sexual recidivism rates from the Static-99R alone, however, the data analysis reveals significant problems in detecting the differences because of the reliance on observed rates of sexual reoffense and small cell frequencies. The determination about whether the Static-99R/ STABLE-2007 produce sexual recidivism rates that are greater than that predicted by the Static-99R alone awaits further research that addresses the limitations of the exploratory study. Until then, evaluators would be hard pressed to reasonably rely upon the standardized AAA to support a qualitative conclusion that the likelihood of sexual reoffense based on the combined measures (without providing a specified probability) is presumed greater than the rate predicted by the sexual violence risk actuarial measure alone.

The variability in the results across studies examining the incremental predictive validity of the STABLE-2007 over the Static-99R, as presented in Table 2, does not inform whether the differences resulted from chance factors (within study sampling error) as opposed to true variability across samples. Meta-analytic research would help to address this conundrum. Brankley et al. (2019) conducted a meta-analysis of twelve studies that employed the Static/STABLE combination, including the eight of the studies listed in Table 2. The remaining studies consisted of samples where the Static/STABLE were administered with persons who had been living in the community. One of the community samples comprised 62% of the total

aggregate sample size for the twelve studies. Brankley et al. (2019) discovered that the variability in measuring the effect of the standardized AAA across studies was in the range expected by sampling error. It is uncertain as to what extent this finding would hold if only the eight studies listed in Table 2 were subject the meta-analysis. Even if the results from Brankley et al. (2019) were applicable to the eight in-custody samples, it does not address the most relevant evidence to evaluate the SVP likely threshold- does the probability of sexual reoffense predicted by the combined instruments is significantly greater than the rate predicted by the Static-99R alone?

The results from the exploratory study illustrate the need to conduct higher powered statistical analysis to determine whether the probability of sexual reoffense is greater when considering the Static/STABLE combination than the rate of sexual reoffense as predicted by the Static-99R alone. In the meantime, evaluators should be aware of two limitations of the standardized AAA. One, evaluators lack a reasonable basis to conclude that the Static-99R alone provides insufficient information to reliably evaluate whether the evaluatee presents a likelihood of sexual reoffending commensurate with the SVP likely threshold. Two, experts lack a reasonably reliable basis to assert that the Static/STABLE combination produces a likelihood of sexual reoffense, which is not quantifiable, but it is assumed to be greater than the Static-99R alone and the outcome supports an opinion that the evaluatee meets the SVP likely threshold. The pitfall of this opinion can be avoided by evaluators relying upon the Static/STABLE priority category sexual reoffense probabilities published in the STABLE 2007 evaluator's workbook (Brankley et al., 2017), but it would still be necessary to test whether the probability of sexual recidivism from the joint measures is meaningfully greater than that predicted by the Static-99R alone and

whether the risk data applies to SVP evaluatees. The following provides guidelines for this analysis.

Brankley et al. (2017) report five-year sexual reoffense rates by Static/STABLE priority categories. The forensic evaluator should first identify the appropriate priority category associated with the evaluatee and reference the point estimate and the 95% confidence interval. For instance, the evaluatee is assigned the Static-99R total score of 7 and the STABLE-2007 total score of 12. The combination of Static/STABLE total scores place him in well above average priority category with a 26.8% sexual reoffense rate over five years and a 95% confidence interval between 17.4% and 36.3%. The next step is to determine whether the Static-99R score-wise risk estimate from the selected reference group (i.e., routine corrections or preselected high risk need), in this example at the score of 7, falls within the 95% confidence interval for the Static/STABLE priority category. The Static-99R five-year sexual recidivism rate from the preselected high risk need reference group is 30.7%. While the Static-99R score-wise point estimate is greater than that predicted by the Static/STABLE priority category, the difference does not appear significant since the Static-99R point estimate falls within the 95% confidence interval for the STABLE/Static well above average priority category. The same outcome happens if the forensic evaluator selected the routine corrections reference group, where the risk estimate is 27.2% at the Static-99R total score of 7. The theory of the standardized AAA is contradicted when there is no significant difference in the rates of sexual violence risk as determined by the Static/STABLE priority category and the Static-99R alone. There are some situations where this comparison will reveal estimates of sexual reoffense determined by the Static/STABLE priority category that are significantly greater than the rates predicted by the Static-99R alone. This situation highlights the need for evaluators to provide triers of fact

objective evidence that the structured AAA applied with evaluatees generate higher probabilities of sexual reoffense than the rates determined by the sexual violence risk actuarial measure alone.

The results from Static/STABLE sexual recidivism actuarial table (Brankley et al., 2017) presume that the individual being assessed is fungible with the sample from which the actuarial data was generated. When this assumption is satisfied, the forensic practitioner has confidence that the point estimate from the actuarial table is the best approximation of the likelihood of sexual reoffense for the individual being assessed (Woodworth & Kadane, 2004). It appears questionable, in my opinion, whether individuals petitioned for civil confinement as SVP are fungible with a sample comprised of individuals from Canada who are being supervised in the community under probation or parole as reported by Brankley et al. (2017). This raises reasonable doubts about the accuracy of the probabilities for sexual reoffense reported in the Static/STABLE actuarial table (Brankley et al., 2017) as applied to individuals undergoing legal proceeding for civil confinement as SVP. Therefore, it would be appropriate for forensic practitioners who rely upon this data for opinions to make known this limitation regarding the validity (American Psychological Association, 2013).

### **Conclusions and Recommendations for Forensic Practice and Research**

Scientific evidence lacks to support the application of the unstructured AAA in SVP civil commitment forensic evaluations. Evaluators who use the unstructured AAA are obligated to appropriately qualify its limitations when rendering opinions in reports and when testifying in legal proceedings (American Psychological Association, 2013; American Psychological Association, 2010). It is imperative, in my view, that evaluators reveal that the unstructured AAA does not produce a quantifiable likelihood of sexual reoffense that is necessary to reliably evaluate whether the evaluatee meets the legally defined likely threshold. Moreover, the trier of

fact should be made aware that this procedure lacks standardized and valid rating criteria for selected risk items, and the method has unknown reliability. It seems inconceivable that unstructured AAA in its present state would offer the trier of fact sufficiently reliable, relevant, and probative evidence that evaluatees meet the SVP likely standard as a result of suffering from an SVP mental disorder.

The limitations and problems of the unstructured AAA as applied to SVP risk assessments could be rectified with substantial research efforts. For example, various DRF could be selected and standardized rating criteria developed for each. The standardized rating criteria would need to be subject to interrater reliability studies and, depending on the outcome, the revision of the rating criteria may be necessary. Once reliable DRF rating criteria are established, the DRF and sexual violence risk measures could be analyzed using the multivariate Bayesian classification method as used by Mokros et al. (2010) or multivariate logistic regression models. This would permit developing a prediction model with the optimal number of sexual violence risk instruments and DRF that maximize predictive accuracy. The prediction model would need to be validated in a large sample of sexual offenders, including the determination of predicted rates of sexual reoffense. The prediction model would then need to be replicated across other samples of sexual offenders.

The standardized AAA shows greater promise than the unstructured AAA for application in SVP risk assessments based on the state of the research in his area. As demonstrated previously, it appears that the Static/STABLE combination incrementally predicts an increased hazard of sexual recidivism beyond the Static-99R alone in samples of men who were in custody at the time of the STABLE-2007 assessment. What is uncertain; however, is whether incremental predictive validity translates into probabilities of sexual reoffense for the combined

measures that are materially different than the Static-99R score-wise probability for sexual reoffense alone. Because research has not examined this critical issue, I conducted the previously described exploratory study with the results reported in Tables 3 and 4. While the results from the exploratory study suggest that the associated point estimates across all Static/STABLE priority categories were not significantly different than the associated Static-99R score-wise sexual reoffense rates, the findings may have been an artifact of methodological problems of relying upon observed sexual recidivism rates with some small cell frequencies. Existing research has been conducted on different models of the standardized AAA, including the Static/STABLE (Brankley et al., 2017) and Static-99R/VRS-SO (Olver et al., 2018) that have produced predicted sexual recidivism rates based on Cox regression survival analysis. This research could be expanded by incorporating additional analyses to test the extent to which the probabilities of sexual recidivism from the combined measures are significantly greater than the rates predicted by the actuarial measure alone.

At this point, the extant research related to the Static/STABLE combination does not, in my view, appear to support the standardized AAA premise that the consideration of the Static/STABLE priority categories uniformly increase the probability of sexual reoffense as predicted by the Static-99R alone. Comparison of five-year sexual reoffense data reported by Brankley et al. (2017) to the Static-99R actuarial tables for the routine corrections or preselected high risk needs reference groups ([http://www.static99.org/pdfdocs/Supplemental\\_Recidivism\\_Tables\\_Static-99R\\_Static-2002R.pdf](http://www.static99.org/pdfdocs/Supplemental_Recidivism_Tables_Static-99R_Static-2002R.pdf)) indicate that the predicted sexual reoffense rate for certain priority categories are less than or no different than that predicted by the Static-99R alone and in some score combinations the sexual recidivism rates are greater than predicted by the Static-99R alone. Since the researchers did not test for differences in predicted sexual reoffense

rates between the Static/STABLE priority categories and the associated Static-99R total score, it is incumbent upon evaluators to make this comparison using the procedure described earlier so as not to mislead the trier of fact in situations where the DRF do not produce a significant increase in the rate of sexual reoffense over the actuarial instrument. Last, but not least, the characteristics of the DSP sample appear sufficiently dissimilar to the SVP population to satisfy the requirement of mutual exchangeability and this raises the related question about the accuracy of the risk data when rendering opinions about the sexual reoffense potential of individuals undergoing SVP civil confinement proceedings. This would not preclude evaluators from relying upon the Static/STABLE priority risk estimates as long as the opinion is appropriately qualified by the limits of generalizability of the data.

The issue about the fungibility of the individual being assessed in SVP risk assessments with the actuarial sample may be avoided when using the Static-99R and VRS-SO dynamic risk measures (“Static/VRS-SO”). The Static/VRS-SO actuarial data is comprised of an aggregate of 913 subjects from four nonoverlapping samples of treated sexual offenders (Olver et al., 2018). An Excel workbook calculator is available at [http://www.psynergy.ca/VRS\\_VRS-SO.html](http://www.psynergy.ca/VRS_VRS-SO.html) that generates predicted estimates of sexual reoffense over five and ten years after inputting the total scores from the Static-99R and VRS-SO dynamic risk measure. The study did not test the extent to which the predicted rates from the Static/VRS-SO differed significantly from the associated Static-99R score-wise sexual recidivism rate from the routine corrections or preselected high risk needs reference groups. This comparison is necessary to avoid misleading the trier of fact that the Static-99R result alone is insufficient to support an opinion that the individual meets the SVP likely threshold, but the predicted probability produced by the Static/VRS-SO supports this

conclusion. Forensic practitioners can easily conduct a rule of thumb comparison of these predicted rates as described next.

Enter the Static-99R and VRS-SO dynamic instrument total scores, and the appropriate change score (see Olver et al., 2018 for instruction) into the Excel spreadsheet to generate the predicted probability of sexual violence risk for the Static/VRS-SO and note the associated two-tailed 95% confidence interval. Using the same Static-99R total score entered into the Excel spreadsheet, obtain the predicted score-wise probability of sexual reoffense from one of the two Static-99R reference groups (Phenix et al., 2016). If the score-wise sexual reoffense rate from the selected Static-99R reference group falls within the bounds of the 95% confidence interval for the predicted Static/VRS-SO predicted rate, then it is unlikely that the two values differ significantly. For a more exact analysis of the differences in sexual recidivism rates, readers can apply the method presented by Cumming and Finch (2005) for comparing results from two independent groups.

It is common for individuals being petitioned for civil commitment as SVP not to participate in sexual offense treatment. When this situation is present at the time of evaluation, it is uncertain whether individuals undergoing SVP evaluations are fungible with members of the Static/VRS-SO actuarial class, which consists of individuals who were treated in prison based treatment programs with varying levels of intensity (Olver et al., 2018). Meta-analytic studies of sexual offender treatment program indicate the base rate of sexual reoffense for treated sexual offenders is 30% to 40% lower than untreated sexual offenders, with the real differences ranging between 4% to 8% (Hanson et al., 2002; Losel & Schmucker, 2005; Schmucker & Losel, 2015; Gannon et al., 2019). Such differences in sexual recidivism base rates and unknown true variability between samples on risk relevant characteristics raise a legitimate question whether

the sexual recidivism estimates from the Static/VRS-SO are generalizable to individuals being petitioned for civil commitment as SVP who are not participating in treatment. Olver et al. (2018) propose a solution for dealing with this situation by recommending a specific change score for individuals who are not participating in sexual offender treatment. Suffice it to say that the validity of this recommendation has not been tested scientifically and it is beyond the scope of this article to address it in detail. If individuals undergoing evaluations for judicial commitment as SVP are involved in sexual offender treatment, then it would appear to be justified to use the Static/VRS-SO.

## REFERENCES

- Abbott, B. R. (2017). Sexually violent predator risk assessments with the Violence Risk Appraisal Guide-Revised: A shaky practice. *International Journal of Law and Psychiatry*, 52, 62-73. <https://doi.org/10.1016/j.ijlp.2017.03.003>
- Abbott, B. R. (2011). Throwing the baby out with the bathwater: Is it time for clinical judgment to supplement actuarial risk assessment? *The Journal of the American Academy of Psychiatry and the Law*, 39(2), 222-230.
- Abracen, J & Looman, J. (2015). *Treatment of High-Risk Sexual Offenders: An Integrated Approach*. Wiley-Blackwell.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- American Psychological Association. (2013). *Specialty guidelines for forensic psychology*. Washington, D.C.: American Psychological Association.
- American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct*. Washington, D.C.: American Psychological Association.
- Association for the Treatment of Sexual Abusers. (2014). *Practice guidelines for the assessment, treatment, and management of male adult sexual abusers*. Beaverton, OR: Association for the Treatment of Sexual Abusers.
- Boccaccini, M. T., Murrie, D. C., Caperton, J. D., & Hawes, S. W. (2009). Field validity of the Static-99 and MnSOST-R among sex offenders evaluated for civil commitment as sexually violent predators. *Psychology, Public Policy, and Law*, 15(4), 278-314. <https://psycnet.apa.org/doi/10.1037/a0017232>

- Boccaccini MT, Murrie DC, Mercado C, et al. Implications of Static-99 Field Reliability Findings for Score Use and Reporting. *Criminal Justice and Behavior*. 2012;39(1):42-58. <https://doi.org/10.1177/0093854811427131>
- Brankley, A. E., Helmus, L. M., & Hanson, R. K. (2017, May). *STABLE-2007 evaluator handbook revised 2017*. Unpublished document.
- Brankley, A. E., Babchishin, K. M., & Hanson, R. K. (2019). STABLE-2007 demonstrates predictive and incremental validity in assessing risk relevant propensities for sexual offending: A meta-analysis. *Sexual Abuse*, 1-29. <https://doi.org/10.1177%2F1079063219871572>
- Christie, R. and Geis, F. L. (1970) *Studies in Machiavellianism*. New York: Academic Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cook, N. R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, 115, 928–935. <https://doi.org/10.1161/CIRCULATIONAHA.106.672402>
- Cumming, G. & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, 60(2), 170-180. <https://doi.org/10.1037/0003-066X.60.2.170>
- Duwe, G. & Kim, K. (2016). *The effective use of risk and needs assessment instruments in the management of juveniles who having sexually offended*. Presentation at the 35th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Orlando, FL. November 3, 2016.

- Duwe, G. & Rocque, M. (2018). The home field advantage and the perils of professional judgment: Evaluating the performance of the Static-99R and the MnSOST-3 in predicting sexual recidivism. *Law and Human Behavior*, 42(3), 269-279.  
<https://psycnet.apa.org/doi/10.1037/lhb0000277>
- Eher, R., Olver, M. E., Heurix, I., Schilling, F., & Rettenberger, M. (2015). Predicting reoffense in pedophilic child molesters by clinical diagnoses and risk assessment predicting reoffense in pedophilic child molesters by clinical diagnoses and risk assessment. *Law and Human Behavior*, 39, 571-580. <https://doi.org/10.1037/lhb0000144>
- Eher, R., Rettenberger, M., Gaunersdorfer, K., Haubner-MacLean, T., Matthes, A., Schilling, F., & Mokros, A. (2013). Ober die treffsicherheit der standardisierten risikoeinschätzungsverfahren Static-99 und STABLE-2007 bei aus einer sicherungsmaßnahme entlassenen sexualstraftatern [On the accuracy of the standardized risk assessment procedures Static-99 and STABLE-2007 for sexual offenders released from detention]. *Forensische Psychiatrie, Psychologie, Kriminologie*, 7, 264-272. <https://doi.org/10.1007/s11757-013-0212-9>
- Eher, R., Rettenberger, M., Matthes, A., Schilling, F. (2012). Dynamic Risk Assessment in Sexual Offenders Using STABLE-2000 and the STABLE-2007: An Investigation of Predictive and Incremental Validity. *Sexual Abuse*, 24(1), 5-28.  
<https://doi.org/10.1177%2F1079063211403164>
- Etzler, S., Eber, R., & Rettenberger, M. (2018). Dynamic risk assessment of sexual offenders: Validity and dimensional structure of the STABLE-2007. *Assessment*. First published online February 6, 2018. <https://doi.org/10.1177/1073191118754705>

Fernandez, Y., Harris, A. J. R., Hanson, R. K., & Sparks, J. (2014, October). *STABLE-2007 Coding Manual Revised 2014*. Unpublished manual.

Gannon, T. A., Olver, M. E., Mallion, J. S., & James, M. (2019). Does specialized psychological treatment for offending reduce recidivism? A meta-analysis examining staff program variables as predictors of treatment effectiveness. *Clinical Psychology Review*, 73, 1-18. <https://doi.org/10.1016/j.cpr.2019.101752>

Hanson, R. K. & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*. 66(2), 348-362. <https://psycnet.apa.org/doi/10.1037/0022-006X.66.2.348>

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*, 14(2), 169-194. <https://doi.org/10.1177%2F107906320201400207>

Hanson, R. K., Harris, A. J. R., Scott, T. L., & Helmus, L. (2007). *Assessing risk of sexual offenders on community supervision: The dynamic supervision project*. Retrieved March 23, 2008, from [http://www.publicsafety.gc.ca/res/cor/rep/\\_fl/crp2007-05-en.pdf](http://www.publicsafety.gc.ca/res/cor/rep/_fl/crp2007-05-en.pdf)

Hanson, R. K. & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders : A meta-analysis of 118 prediction studies. *Psychological Assessment*, 21(1), 1-21. <https://doi.apa.org/doi/10.1037/a0014421>

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

- Harris, G. T. & Rice, M. E. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior*, 29(5), 615-620. <https://doi.org/10.1007/s10979-005-6832-7>
- Harris, G. T., Rice, M. E., Quinsey, V. L., & Cormier, C.A. (2015). *Violent Offenders: Appraising and Managing Risk*. Washington, DC: American Psychological Association.
- Helmus, L. M., & Hanson, R. K. (2013). *STABLE-2007: Updated recidivism rates (includes combinations with Static-99R, Static-2002R, and Risk Matrix 2000)*. Unpublished report. Ottawa, ON: Public Safety Canada.
- Helmus, L., Hanson, R. K., Thornton, D. Babchishin, K. M., & Harris, A. J. R. (2012) Absolute Recidivism Rates Predicted by Static-99R and Static-2002R Sex Offender Risk Assessment Tools Vary Across Samples: A Meta-Analysis, *Criminal Justice and Behavior*, 39, 1148-1171. <https://doi.org/10.1177%2F0093854812443648>
- Jackson, R. L., & Hess, D. T. (2007). Evaluations for civil commitment of sex offenders: A survey of experts. *Sexual Abuse*, 19, 425-448. <https://doi.org/10.1177/107906320701900407>
- Jackson, R., Travia, T., & Schneider, J. (2008). *Annual survey of sex offender civil commitment programs*. *Sexual Offender Civil Commitment Network Research Committee*. Retrieved October 25, 2009, from: [www.soccpn.org/ATSA%202008%20presentations/2008%20Annual%20Survey%20of%20Programs.pdf](http://www.soccpn.org/ATSA%202008%20presentations/2008%20Annual%20Survey%20of%20Programs.pdf).
- 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.

Knight, R. A. & Thornton, D. (2007). *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders*. Washington, D.C.: U.S. Department of Justice.

Knighton, J. C., Murrie, D. C., Boccaccini, M. T., & Turner, D. B. (2014). How likely is “likely to reoffend” in sex offender civil commitment trials? *Law and Human Behavior*, 38(4), 293-304. <https://doi.org/10.1037/lhb0000079>

Looman, J. & Abracen, J. (2014). *Incremental Validity of the STABLE-2007*. Poster presented at the 33rd Annual ATSA Conference San Diego.

Looman, J., & Abracen, J. (2012, October). *Long-term follow-up of two groups of sex offenders*. Presentation at the 31st Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, *Denver, CO*

Looman, J., & Goldstein, J. (2015, October). *Incremental validity of the STABLE-2007 with the Static-99R*. Presentation at the 34th Annual Research and Treatment Conference for the Treatment of Sexual Abusers, *Montreal, QC*

Losel, F. & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology*, 1, 117-146. <https://doi.org/10.1007/s11292-004-6466-7>

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Mann, R. E., Hanson R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse*, 22, 191-217. <https://doi.org/10.1177/1079063210366039>

- Mokros, A., Stadtland, C., Osterheider, M., & Nedopil, N. (2010). Assessment of risk for violent recidivism through multivariate Bayesian classification. *Psychology, Public Policy, and Law*, 16(4), 418-450. <https://doi.apa.org/doi/10.1037/a0021312>
- Murrie, D. C., Boccaccini, M. T., Turner, D., Meeks, M., Woods, C., & Tussey, C. (2009) Rater (dis)agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and the Law*, 15, 19–53. <https://doi.apa.org/doi/10.1037/a0014897>
- Olver, M. E., Mundt, J. C., Thornton, D., Beggs, S.M., Kingston, D. A., Sowden, J. N., Nicholaichuk, T. P., Gordon, A., & Wong, S. C. P. (2018). Using the violence risk scale-sexual offense version in sexual violence risk assessments: Updated risk categories and recidivism estimates from a multisite sample of treatment offenders. *Psychological Assessment*, 30(7), 941-955. <https://doi.org/10.1037/pas0000538>
- Olver, M. E., Stockdale, K. C., Neumann, C. S., Hare, R. D., Mokros, A., Baskin-Sommers, A., Brand, E., Folino, J., Gacono, C., Gray, N. S., Kiehl, K., Knight, R., Leon-Mayer, E., Logan, M., Meloy, J. R., Roy, S., Salekin, R. T., Snowden, R., Thomson, N., . . . Yoon, D. (2020). Reliability and validity of the Psychopathy Checklist-Revised in the assessment of risk for institutional violence: A cautionary note on DeMatteo et al. (2020). *Psychology, Public Policy, and Law*, 26(4), 490–510. <https://doi.org/10.1037/law0000256>
- Phenix, A., Fernandez, Y., Harris, A. J. R., Helmus, M., Hanson, R. K., & Thornton, D. (2016). *Static-99R coding rules- revised 2016*. Available at: [http://www.static99.org/pdffdocs/Coding\\_manual\\_2016\\_v2.pdf](http://www.static99.org/pdffdocs/Coding_manual_2016_v2.pdf)

- Phenix, A., Helmus, L. M., & Hanson, R. K. (2016). *Static-99R and Static-2002R evaluators' workbook*. [https://www.static99.org/pdfdocs/Evaluators\\_Workbook\\_2016-10-19.pdf](https://www.static99.org/pdfdocs/Evaluators_Workbook_2016-10-19.pdf)
- Prentky, R. A., Janus, E., Barbaree, H., Schwartz, B. K., & Kafka, M. P. (2006). Sexually violent predators in the courtroom: Science on trial. *Psychology, Public Policy, and the Law*, 12(4), 357-393. <https://doi.apa.org/doi/10.1037/1076-8971.12.4.357>
- Randolph, J. J. & Edmondson, R. S. (2005). Using binominal effect size display (BESD) to present magnitude of effect sizes for the evaluation audience. *Practical Assessment, research, & Evaluation*, 10(14), 1-7. <https://doi.org/10.7275/zqwr-mx46>
- Salgado, J. F. (2018). Transforming the area under the normal curve (AUC) into Cohen's d, Pearson's  $r_{pb}$ , and natural log odds-ratio: Two conversion tables. *The European Journal of Psychology Applied to Legal Context*, 10(1), 35-47. <https://doi.org/10.5093/ejpalc2018a5>
- Saum, S. (2004). *A comparison of an actuarial risk prediction measure (Static-99) and a STABLE dynamic prediction measure (STABLE-2000) in making risk predictions for a group of sexual offenders* (Unpublished doctoral dissertation). The Fielding Institute, Santa Barbara, CA.
- Schmucker, M. & Losel, F. (2015). The effects of sexual offender treatment of recidivism: An international meta-analysis of sound quality evaluations. *Journal of Experimental Criminology*, 11(4), 597-630. <https://doi.org/10.1007/s11292-015-9241-z>
- Schneider, J. E., Jackson, R., D'Orazio, D., Hebert, J., & McCulloch, D. (2014, October). *SOCPCN annual survey of sex offender civil commitment programs 2014*. Paper presented at the 2014 SOCPCN Annual Meeting, San Diego, CA. Retrieved January 5, 2015, from [http://socpcn.org/images/SOCPCN\\_Annual\\_Survey\\_2014\\_revised.pdf](http://socpcn.org/images/SOCPCN_Annual_Survey_2014_revised.pdf)

Scurich, N. & Krauss, D. (2014). The presumption of dangerousness in sexual violent predator commitment hearings. *Law, Probability, and Risk*, 13, 91-104.

<https://doi.org/10.1093/lpr/mgt015>

Scurich, N. & Krauss, D. (2013). The effect of adjusted actuarial risk assessment on mock-jurors' decisions in sexual predator commitment proceedings, *Jurimetrics*, 53, 395-413.

Sowden, J. N., & Olver, M. E. (2017). Use of the Violence Risk Scale-Sexual Offender version and the STABLE-2007 to assess dynamic sexual violence risk in a sample of treated sexual offenders. *Psychological Assessment*, 29, 293-303.

<https://doi.org/10.1037/pas0000345>

Sreenivasan, S., Weinberger, L. E., Frances, A., & Cusworth-Walker, S. (2010). Alice in actuarial -land: Through the looking glass of changing Static-99 norms. *Journal of the American Academy of Psychiatry and the Law*, 38, 400-406.

Streiner, D. L. & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *Canadian Journal of Psychiatry*, 52(2), 121-128.

<https://doi.org/10.1177/070674370705200210>

Storey, J. E., Watt, K. A., Jackson, K. J., & Hart, S. D. (2012). Utilization and implications of the Static-99 in practice. *Sexual Abuse*, 24(3), 289-302.

<https://doi.org/10.1177%2F1079063211423943>

Thornton, D. (2003). The Machiavellian sex offender. In A. Matravers (Ed.), *Sex offenders in the community: Managing and reducing the risks* (pp. 144-152). Cullompton, England: Willan.

- Thornton, D. and Hogue, T. (1993) 'The Large Scale Provision of Programmes for Imprisoned Sex Offenders: issues, dilemmas and progress, *Criminal Behaviour and Mental Health*, 3(4), 371-80. [https://doi.org/ 10.1002/cbm.1993.3.4.371](https://doi.org/10.1002/cbm.1993.3.4.371)
- Van den Berg, J. W., Smid, W., Schepers, K., Wever, E., van Beek, S., Janssen, E., & Gijs, L. (2018). The predictive properties of dynamic sex offender risk assessment instruments: A meta-analysis, *Psychological Assessment*, 30(2), 179-191. <https://doi.org/10.1037/pas0000454>
- Vrieze, S. L. & Grove, W. M. (2010). Multidimensional assessment of criminal recidivism: Problems, pitfalls, and proposed solutions. *Psychological Assessment*, 22(2), 382-395. <https://doi.apa.org/doi/10.1037/a0019228>
- Ward, T., & Beech, A. R. (2015). Dynamic risk factors: A theoretical dead-end? *Psychology, Crime & Law*, 21, 100–113. <https://doi.org/10.1080/1068316X.2014.917854>
- Woodworth, G. G. & Kadane, J. B. (2004). Expert testimony supports post-sentence civil incarceration of violent sexual offenders. *Law, probability, and risk*, 3, 221-241. <https://doi.org/10.1093/lawprj/3.3-4.221>
- Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior*, 39, 1511–1538. <https://doi.org/10.1177/0093854812455741>