

What Can We Reasonably Expect to Accomplish in Conducting Actuarial Risk Assessments with Sexual Offenders in Civil Commitment Settings? A Response to Campbell and DeClue: “Maximizing Predictive Accuracy in Sexually Violent Predator Evaluations”

Authors: Robin J. Wilson & Jan Looman

Robin J. Wilson, Ph.D., ABPP, Florida Civil Commitment Center, Arcadia, Florida
dr.wilsonrj@verizon.net

Jan Looman, Ph.D., Regional Treatment Centre, Correctional Service Canada,
Kingston, Ontario

Abstract

Measuring risk to recidivate in populations of sexual offenders has always been a contentious and complicated process. Whereas predictions of reoffense risk were once offered by clinicians using “unstructured clinical judgment,” the research literature has clearly shown that the accuracy of such methods often fails to meet even chance levels. Consequently, the past 20 years or more has seen the development and refinement of tools designed to actuarially predict risk for recidivism. One such tool, the Static-99, has become the index of choice for most evaluators of sexual recidivism risk, to the extent that it is likely the most widely researched and utilized measure of its type in the world. However, the Static-99 is not without its foibles: It possesses only moderate predictive accuracy, omits such robust predictors as direct measures of sexual deviancy and antisociality, and does not include any dynamic factors. Further, it appears that heterogeneity in the standardization samples presents issues for evaluators and jurists alike in terms of how best to evaluate individual offenders against putative normative groups. In their paper “Maximizing Predictive Accuracy in Sexually Violent Predator Evaluations,” Campbell and DeClue argue that the predictive value of the Static-99 and its related measures is suspect. These arguments center on issues related to cut-off scores, confidence intervals, and assignment of individual offenders to specific normative samples based on factors not included in the Static-99. In this paper, we provide a response to some of the issues raised by Campbell and DeClue, specifically focusing on forcing dichotomous status on continuous variables, a failure to see the big picture in using Static-99 to predict risk in populations of Sexually Violent Predators, and difficulties associated with creating confidence intervals.

Key words: Static-99, sexual recidivism, actuarial prediction, sexually violent predator, civil commitment

Wilson, R. J., & Looman, J. (2010). What can we reasonably expect to accomplish in conducting actuarial risk assessments with sexual offenders in civil commitment settings? A response to Campbell and DeClue: “Maximizing predictive accuracy in sexually violent predator evaluations. *Open Access Journal of Forensic Psychology*, 2, 306-321.

The assessment and management of risk to reoffend in criminal offenders is a matter of considerable public and professional interest. As society becomes more and more intolerant of violence in its midst, legislators have moved to impose measures intended to lower the potential for harm caused to the public by known offenders. This is especially true in regard to sexual offending, where recent years have seen the implementation of a number of policies designed specifically to reduce sexual recidivism. Although some researchers have challenged many of these measures as being high in public appeal but low in efficacy (see Levenson & D'Amora, 2007), it is not our intention to address that particular issue in this paper.

As public policies have been implemented in an attempt to manage sexual recidivism from an “official control” standpoint, behavioral scientists have attempted to develop methods to better identify who may or may not reoffend once released back to the community. Indeed, as pertains to sexual offender risk assessment, this has become a contentious issue for empirical and practical debate. At the heart of the matter is whether or not risk to reoffend can be accurately predicted. This issue is particularly pronounced in the United States, where 20 states and the federal government have enacted procedures for the indeterminate civil commitment of persons judged to be “sexually violent predators” or “sexually violent persons,” known colloquially as SVPs. The utility of statutory definitions associated with labeling an offender an SVP are beyond the scope of this paper; however, readers should know that general criteria for this designation are: 1) a history of sexual offending, and 2) possession of a “mental abnormality” that, without appropriate treatment, would preclude the individual from being able to control his/her propensity to engage in further sexual offending. A key component to establishing the second criteria typically involves assessments of the offender’s risk to reoffend.

In “Maximizing Predictive Accuracy in Sexually Violent Predator Evaluations”, Campbell and DeClue (2010) argue that there are many issues that evaluators must consider should they choose to use actuarial risk-assessment instruments in their evaluations of persons referred for possible civil commitment as an SVP. In the following, we offer a response to some of their premises and concerns.

Actuarial Risk Assessment

Over the past three decades research scientists have worked hard to refine methods of assessing risk; particularly to address concerns raised by Monahan (1981) that unstructured clinical judgment regarding risk to reoffend led to accuracy levels below chance (i.e., we would have been better off flipping a coin than asking a risk-assessment professional for his/her opinion). Early efforts to reduce subjectivity in the risk-assessment process included the Level of Service Inventory (see LSI-R—Andrews & Bonta, 2000), the General Statistical Information on Recidivism scale (GSIR—Nuffield, 1982; later revised and known as the SIR-R1—Nafekh & Motiuk, 2002), and the Violence Prediction Scheme (an early precursor to the Violence Risk Appraisal Guide [VRAG] and the Sexual Offender Risk Appraisal Guide [SORAG]—see Quinsey, Harris, Rice, & Cormier, 2006).

Actuarial risk assessment involves the use of scales comprised of empirically derived factors with some demonstrated utility in predicting a certain type of behavior—in this case sexual or violent recidivism. The items are often chosen through a process of meta-analysis of factors (e.g., Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005). Standardization of the measure requires scoring a sample of individuals on the scale, following them for a period while they are exposed to risk, and then statistically analyzing the outcome data. At the completion of the time-at-risk period, the scale's constructors will calculate what percentage of persons in each cohort of scores has reoffended (e.g., 35% of persons who scored "5" on the scale reoffended within a period of 10 years). It is then possible to use the instrument on an individual knowing that, within a margin of error, an individual who scores "5" is likely to behave similarly to the larger group of persons with that same score in the standardization sample. We might then say, "Research has shown that groups of men demonstrating the same empirical risk factors as Mr. X have been seen to recidivate at 35% over 10 years."

Static-99

The Static-99 (Hanson & Thornton, 1999) and related instruments (Static-99R, Static-2002, Static-2002R) are among the most popular and extensively researched tools of their kind in the world (Helmus, Thornton, & Hanson, 2009). The Static-99 actuarially assesses risk for sexual and violent recidivism based on static risk variables (i.e., variables that are largely historical in nature and not amenable to intentional change or clinical influence). The original 1999 version consisted of 10 static items and scores ranged from 0 to 12. Moderate to good predictive validity has been found for the Static-99 (average $d = .67$ across 63 studies—Hanson & Morton-Bourgon, 2009). In 2008, new recidivism norms were released following concern that the 1999 recidivism estimates over-predicted risk, especially at the higher score levels (Harris, Helmus, Hanson, & Thornton, 2008). In 2009, one item of the original scale (age at release) was changed to allow for a greater range of scoring opportunities consistent with contemporary research showing that risk decreases as offenders age (Helmus et al., 2009). Scores on the Static-99R range from -3 to 12 (Helmus, 2009). Recidivism estimates were again revised and released, now allowing evaluators to compare individual cases against four normative samples. However, the criteria for choosing against which sample to compare individual offenders has continued to be a subject of much discussion.

Since its release, the Static-99 and its related instruments (Static-99R, Static-2002, Static-2002R) have been the subject of considerable cross-validation research in many jurisdictions throughout North America, the United Kingdom, Europe, and Asia. For the most part, predictive accuracy in those jurisdictions has been similar, as was outlined in a recent meta-analysis (Hanson & Morton-Bourgon, 2009). However, in spite of its broad popularity and consistently demonstrated empirical utility, the Static-99 and its progeny are not without their foibles, which are arguably applicable to all such scales (e.g., SORAG, MnSOST-R, etc.).

Problems in Actuarial Risk Assessment with SVPs

Before we turn our attention to problems specific to measuring risk potential in SVPs, let us first acknowledge that there are problems inherent in actuarial risk assessment as *applied to all populations*. Studies of the accuracy of actuarial risk prediction tools show clearly that such measures have typically moderate predictive ability (e.g., for Static-99: $d = .67$, with $N = 20,010$ over 63 studies contributing to the meta-analysis—Hanson & Morton-Bourgon, 2009). In their paper, Campbell and DeClue acknowledge that construction of a tool that includes all sources of predictive ability would be impractical. Indeed, the seminal Hanson meta-analyses (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005)—the earlier of which had a major impact on item selection for Static-99—showed clearly that there are many factors that might assist us in predicting risk. However, inclusion of all such items would result in an unwieldy tool, unpopular with assessors and of limited utility in some jurisdictions. Regarding the latter, penile plethysmography (also known as phallometry—see Freund, 1963) is a case in point. Both of the aforementioned Hanson meta-analyses identified possession of deviant sexual interests (via sexual preference testing) as being the most robust predictor of sexual recidivism, yet the Static-99 and its related scales do not include this as an item. This is partly because not all jurisdictions utilize sexual preference testing methods; sometimes because local social mores do not support use of testing that relies on presentation of sexually deviant stimuli (pictures, videos, or narratives). Further, engagement in sexually deviant behavior appears to be amenable to change via treatment (Hanson, Bourgon, Helmus, & Hodgson, 2009), which suggests that this is a dynamic predictor. Nonetheless, it is important to note that, from the outset, Static-99 does not include the item with the single greatest predictive power. The implication is that Static-99 is not a fully comprehensive risk assessment, nor has it ever attempted to be.

Sexually violent predators (or persons) are unique among sexual offenders. To put a fine point on this, let us use Florida as an example. Florida's civil commitment statute was enacted in January 1999. The Department of Children and Families' Multi-Disciplinary Team has screened more than 38,699 sexual offenders identified in Florida over the past 11 years (Florida Department of Children and Families flowchart, personal communication with Suzonne Kline, September 2010). However, of those 38,699 screened offenders, approximately nine percent (3,640) were referred for psychological/psychiatric evaluation, and only 1,344 sexual offenders have been referred to the Court for civil trial, which represents about 3.5 percent of the sexual offenders reviewed in Florida since the statute was enacted. Even fewer of those referred for commitment were actually found to be sexual violent predators—in fact, less than half. As it makes some sense that these civil proceedings would be used with only those offenders who pose the greatest risk to the public, the logical supposition is that those who pose the greatest risk are exceedingly rare.

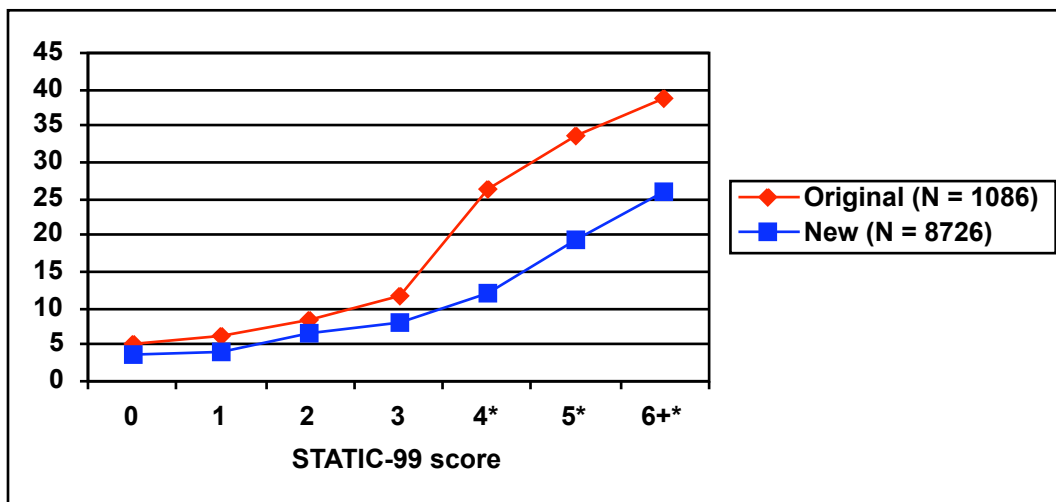
It is commonly known that the general distribution of risk to reoffend among sexual offenders is positively skewed, meaning that there are many more offenders on the lower end of the risk continuum than on the higher end (see Hanson & Thornton, 1999).

Indeed, in the 2009 data (Phenix, Helmus, & Hanson, 2009) for the Static-99R standardization sample, 34.9 percent of the included offenders scored “1” or less (nominal risk category of “low”) while only 15.2 percent of the sample scored “6” or higher (maximum score of 12), described by the test’s constructors as being in the nominal category “high risk.”

At this point, it is important to define what we mean by “high risk.” Many SVP statutes have adopted language along the lines of “more likely to (recidivate) than not” when discussing the inclusion criteria for who should be considered for civil commitment. The original Static-99 standardization sample appeared to serve this concept well, in that those offenders who scored “6” or higher were nominally assessed as both “high risk” and at a 52 percent chance of reoffending in 15 years (i.e., more likely to than not). From a social perspective, there is a certain degree of risk that the community may abide. When faced with a 52 percent chance of thunderstorms, many of us will still choose not to carry an umbrella. Clearly, we cannot equate the risk of getting wet with the risk of being sexually assaulted—the stakes are in no way equivalent. The risk for not reoffending in the “high-risk” nominal group in the original Static-99 sample is only slightly less than the risk for reoffending (48 percent vs. 52 percent, within margins of error). Indeed, many in the community might ascribe high-risk status to offenders showing even lower risk probabilities than 52 percent.

In 2008 (Hanson & Thornton, 2008), new normative samples delineating lower projected rates of sexual recidivism were released by the Static-99 research group, largely because contemporary research has suggested that, even within the same Static-99 score, sexual offenders are reoffending at rates lower than those reported when the original Static-99 normative sample was established (as well-demonstrated in Figure 1, below). In fact, in some cases, researchers have found rates to be dramatically lower than projected, even in high-risk samples (e.g., Wilson, Picheca, & Prinzo, 2007).

Figure 1:
Comparisons of percent recidivism rates by score—Five-year Norms
(from Helmus, 2009; reprinted with permission)



When preliminary data were presented in the fall of 2008, the Static-99 research group suggested that samples could be divided in two—"routine" cases versus those who would be considered "preselected for high risk." However, within a year, when the project was completed, there was a new formulation for proposed normative samples. At the 2009 conference of the Association for the Treatment of Sexual Abusers (ATSA) the Static-99R was introduced, along with sets of norms for samples with different base-rates, even within the same Static-99 score (i.e., samples described as "routine", "non-routine", "preselected for treatment", and "preselected for high risk/needs"—Helmus et al., 2009; Hanson, Phenix, & Helmus, 2009). It is important to note that each of these subsamples has significantly lower base-rates of reoffending than the 1999 norms, even the group preselected for high risk/needs.

Maximizing Predictive Accuracy

Campbell and DeClue suggest that using a cut-off of "6" from the original 1999 sample provides the greatest predictive accuracy when one assumes that all persons scoring "6" or higher will reoffend and that those scoring less will not. However, they go on to make the case that, with the new recidivism norms discussed above, the greatest predictive accuracy is achieved when predicting that no one will re-offend, and that the confidence interval that can be ascribed to individual risk predictions (based on group data) is so broad as to be essentially meaningless. They make this argument based on what, on the face of it, appears to be convincingly simple statistical procedures.

However, as Hanson and Howard (2010) point out, one cannot equate sexual offender (or any other offender) risk assessment with medical outcome risk assessment, as Campbell and DeClue attempt to do (providing yes/no predictions for recidivism). The rationale for the distinction is simple: For medical conditions, the presence of certain diagnostic features (runny nose, cough, lethargy) is associated with the likelihood that the person has a particular condition (common cold). This diagnosis can be proven to a reasonable degree of certainty through medical testing. With sexual offender risk assessment the case is not the same, as pointed out by Hanson and Howard (2010). As they say, "In contrast to diagnoses, risk assessments estimate the likelihood of an event that has not yet happened, and may never happen. They are inherently stochastic, and the future outcome can only be estimated with a certain probability" (p. 276). Thus, equating risk assessment with medical diagnosis, as Campbell and DeClue attempt to do, is patently inappropriate. It is more appropriate, as Hanson and Howard also point out, to equate sexual offender risk assessment to the process which insurance companies use when deciding the cost of insurance. A young male is at higher risk for car accidents; so young males pay more for insurance. An overweight smoker is at higher risk for medical complications than a fit nonsmoker, so overweight smokers pay more for medical insurance. The overweight smoker may never experience the adverse medical conditions for a variety of reasons (including genetics), which are currently poorly understood and not assessed by the insurance company. However, reasonable people understand the greater risk ascribed to that person and accept the higher cost for insurance.

However, no reasonable person would say that a 30-year-old male who is overweight and smokes WILL succumb to cancer in the next 10 years, despite his higher statistical probability. Returning to the Static-99, it would be our contention that this unnatural forcing of a continuous variable into dichotomous status fails to appreciate what the instrument was designed to do, in a similar manner, as the overweight smoker/cancer connection is unreasonable. Simply put, the Static-99 and its progeny were not designed to facilitate answering of the “more likely to than not” question faced by SVP evaluators, *no matter how much they may want them to*. Rather, this instrument was designed to be used on sexual offenders with a range of risk profiles. Further, arbitrarily assuming that all offenders with scores of “6” or more will reoffend seems to omit the obvious reality that we already know that 48 percent (within a margin of error) of such offenders will not reoffend, at least according to the 1999 normative sample.

Perhaps, the major problem in this matter regards how we view success and failure in sexual offenders. Campbell and DeClue appear to argue that we must be able to separate offenders into finite groups—those who will and those who will not reoffend. This assumes that success upon release is a binary concept, based solely on the presence or absence of recidivism. However, recent clinical formulations suggest that success is likely better conceptualized as continuous in nature (Pake, 2010). Currently popular philosophies, such as Old Me/New Me (Haaven, Little, Petre-Miller, 1990), Self-Regulation/Good Lives Model (Yates & Ward, 2008), and the “balanced, self-determined lifestyle” concept of the Life Skills model (Curtiss & Warren, 1973; Wilson & Yates, 2009), all stress that treatment must attend to the whole being, and not just focus on sexually offensive behavior. In such formulations, success is more about lessened personal and social dysfunction, with decreased offending being but one aspect of the positive outcomes enjoyed by those who achieve greater lifestyle balance and facility. This is also more in line with the concept of “harm reduction” (Marlatt, 1998), wherein attention to dynamic risk lessens the chance for harm to others while increasing self-regulation.¹ Indeed, this is the major thrust behind advanced risk-assessment and management frameworks such as Thornton’s (2002) Structured Risk Assessment model or the Hanson, Harris, Scott, and Helmus (2007) Dynamic Supervision model (including Stable-2007 and Acute-2007). All in all, attendance to “in-treatment change” (Abracen & Looman, 2005; Hanson & Harris, 2001)—throughout intensive institutional treatment to community-based (and, perhaps, lifelong) maintenance—is probably a more informative endeavor than simply measuring who did or did not commit a new sexual offense. This is particularly true for SVPs, who typically present lifestyle dysfunction in many more domains than just sexual self-regulation.

Margins of Error

You might ask, “Why are they continually restating *within a margin of error?*” Simply, we keep stating it because it is the truth and this reality needs to be acknowledged.

¹ Regarding harm reduction, it is possible to see some degree of positive outcome even in those offenders who recidivate; particularly, with respect to lesser degrees of offense severity and harm sustained by victims.

There is considerable error inherent in the predictions of the Static-99 and associated measures. In fact, this is true of all such instruments, as discussed above with the insurance example. This is why all good risk-assessment trainers admonish their workshop participants against putting all their eggs in the same basket, as it were. Earlier, we explained that not all predictors could realistically be included in a scale. This is one source of prediction error. Another source derives from the manner in which individual offenders achieve certain scores. For example, we have stated above that Static-99 scores of “6” or higher place offenders in the nominal category of “high risk” relative to other offenders achieving different scores; however, how an individual offender achieves that score may vary. Not all variables on the Static-99 contribute the same amount of variance to the risk-prediction effort. In reviewing the tables provided by Hanson and Bussière (1998)—the data that provided some of the rationale for choosing certain items over others for the Static-99—it is clear that not all items eventually included in the instrument possess the same predictive strength. Last, error may result from a lack of reliable or complete file information used to compute scores, or the rater may simply interpret that data incorrectly. Simply put, there are numerous sources of error with which assessors must contend if they wish to provide useful predictions. This underscores the need for comprehensive methods in risk assessment, beyond relying on actuarial predictions alone.

Hanson et al. (2009) have suggested that there are moderator variables that might affect absolute recidivism rates. They concluded that, among a variety of methodological, individual-level, and systems-level factors, this within-score variability is largely due to two factors—age at release (thus, necessitating revision of the “age” item), and sample type. Regarding the latter, when we look at the highly significant *Q* statistics included in the meta-analyses of Static-99 including all replications ($Q = 129.85$, $p < .001$ —Hanson & Morton-Bourgon, 2009; see also Helmus, 2009), we can clearly see that there is a high degree of variability within some putative normative sample types (e.g., all subjects and routine subjects versus subjects preselected for treatment need or subjects preselected as high risk/needs).

There is a problem, however. How do we decide which normative sample to use? What framework should we employ? Thornton, Hanson, and Helmus (2009) suggest that using the standard model of actuarial assessment for Static-99 would be flawed because it assumes (a) that base rates are constant, and (b) that all offenders possess risk factors in essentially the same way. Further, as many of the items are binary, or do not allow for quantifying levels of harm associated with the particular conduct being measured, there is also potential error present in this regard. Although further work needs to be completed to describe the various subsamples sufficiently to allow matching individual offenders to the specific normative samples provided, there is a certain degree of face validity associated with the process. For example, most evaluators will agree that a person who scores a “3” on the Static-99 and who is also phallometrically judged to be pedophilic warrants greater attention than a person with a similar score who does not show pedophilic sexual arousal (accepting that phallometry presents its own set of issues—see Marshall, 2005). Similarly, a person with Static-99 score of “6” combined with a Psychopathy Checklist-Revised (PCL-R—Hare, 2003)

score of “30” causes more concern than someone with a similar Static-99 score and a PCL-R score of “18” (accepting again that even the PCL-R, so popular in SVP settings, presents issues—interested readers may wish to review the recent debate and controversy in *Psychological Assessment* [Hare & Neumann, 2010; Skeem & Cooke, 2010a,b]).

During the 2009 ATSA conference in Dallas, TX, presentations were made that demonstrated incremental predictive value when psychological risk factors were used to revise relative risk assessments (Hanson & Helmus, 2009; Harkins, Thornton, & Beech, 2009; and Thornton & Knight, 2009). Thornton et al. (2009) point to strikingly similar findings in these three data sets that show that the magnitude of the incremental validity of these ratings was sufficient to reproduce the base-rate effects of the different sample types. Table 1 below, from Thornton et al. (2010), shows absolute recidivism estimates when one factors in scores on Stable-2007 or when one differentiates between “routine” and “selected for risk/need” normative groupings. Interestingly, the pattern is quite similar between these two methods of separating the sample.

Table 1
Relative risk ratings considering additional factors
(from Thornton et al., 2009; reprinted with permission)

Static-99R Score	Three-Year Recidivism		Five-Year Recidivism	
	STABLE=5	STABLE=14	Routine	Selected for Risk/Need
2	3%	7%	5%	12%
5	7%	18%	11%	25%
7	14%	32%	19%	38%

SVP Populations

Earlier, we contended that SVPs are not necessarily representative of sexual offenders in general. In a perfect world, all those persons deemed to be SVPs would be in that group of offenders posing the greatest risk to reoffend sexually. However, personal experience (author RJW) suggests this *may not always* be the case. It would appear that potential for sexual recidivism is not the only major risk issue found (or considered) in this select group of offenders. The Hanson and Morton-Bourgon meta-analysis (2004, specifically Table 1 in that paper) showed that sexual deviancy was the single greatest predictor of sexual recidivism. The next most important factor was “antisocial orientation.” Related to SVP populations, the latter often equates to psychopathy, often characterized using PCL-R. Again, using Florida’s SVP program as an example: Among a sample of 121 treatment participants sequentially assessed for treatment track placement, average Static-99 score was 5.96, average score on the PCL-R was 22.72, and average Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 2000) was 27.61. This represents a high risk/needs population and, as such, we would be

generally inclined to suggest that this is the normative sample to which these men should be compared.

In an upcoming presentation to be made at the annual conference of the Association for the Treatment of Sexual Abusers in Phoenix, AR (Wilson, Looman, Abracen, & Pake, 2010), we will present data comparing residents at the Florida Civil Commitment Center (FCCC) to those at the Regional Treatment Centre (RTC), a maximum security correctional treatment center in Canada. Without wanting to give it all away before the presentation, it is our contention that these two groups of offenders are essentially similar to one another, in terms of both psychological makeup and risk potential. The main difference between these two groups is that offenders in the RTC group will be released to the community at or before the completion of their criminal sanction (i.e., sentence) whereas the FCCC group is released only at the discretion of the Court. It is our hope that, if these two groups truly are more similar than not, the release experiences of the RTC group might help better inform us as to the probabilities for success or failure should members of the FCCC group also be released. We also hope to gain increased knowledge of how treatment participation might affect success or failure in the community for this group of high-risk and high-need offenders.

However, regardless of what we might find in our study comparing FCCC to RTC, and further regardless of how myopic this perspective might be, it is clear to us that the SVP field would like to have its own normative sample designed specifically for SVP populations. Hanson et al. (2009) have clearly stated that the best-case scenario for determining recidivism estimates is to establish “local norms,” a position with which Campbell and DeClue would appear to concur. Hanson et al. also state that a general fallback position would be to use the “routine” norms, stating strongly that use of one of the other normative groups would require justification. In the case of SVP, we suggest that local norms would constitute a sample of released civil committees aggregated from the various programs across the USA. Indeed, this would appear to be a very worthy research project. In the alternative, evaluators might be well advised to look for an analog to such a “local norm.” We would argue (and intend to present supportive findings this fall) that the preselected for high-risk/need normative sample is the one that best approximates the attributes of the typical SVP (again, within a margin of error). This is supported to a large degree by the very process that brought these particular offenders to the point that they were even being evaluated for potential SVP status.

With jurisdictional differences, the various SVP statutes generally include two elements: 1) a history of sexual offending, and 2) possession of a “mental abnormality” that, without appropriate treatment, would preclude the individual from being able to control his/her propensity to engage in further sexual offending. Component 2 would seem to be analogous to the “psychological risk factors” concept highlighted by Thornton et al. (2009, p. 3) as being important in justifying the use of normative samples other than “routine”:

It is only when unusually high levels of psychological risk factors are also present that the levels of risk required for civil commitment can reasonably be

attributed. . . . Prudent evaluators would only make this kind of attribution if they can also see direct evidence of high levels of psychological risk factors.

Add to that the fact that a particularly low number of such persons is ever presented to the court for SVP hearings (e.g., 4% of all adjudicated sexual offenders in Florida), and most readers by now should agree that *these are clearly not routine sexual offenders*. Why then, would we advocate use of routine norms?²

Final Thought

We are concerned that readers of the Campbell and DeClue paper might be unintentionally left with the impression that the original Static-99 scoring and norms, using a cutoff score of “6” projecting a 52 percent chance of reoffending in 15 years, represent the most accurate means of judging who will or will not sexually reoffend upon release to the community, at least using Static-99. There is very little reason not to believe the assertion that reoffense rates have declined over the past decade, given that these findings are not unique to Static-99 projections or specific to sexual violence, for that matter (Federal Bureau of Investigation, 2007; Finkelhor & Jones, 2006; Minnesota Department of Corrections, 2007). So, if you believe that the playing field has changed regarding recidivism rates and that the recent research regarding aging and sexual offending is *bona fide* (see Barbaree, Langton, & Blanchard, 2007; Hanson, 2006), then you simply cannot continue to use the original Static-99 scoring or recidivism estimates. Indeed, to do so would be equivalent to using the WAIS-R knowing that a WAIS-III exists, or continuing to use DSM-IV-TR criteria after the DSM-V is released. We believe such practices to be ethically questionable.

Authors' Notes

The views expressed in this paper are those of the authors alone, and do not necessarily reflect the views of the Florida Civil Commitment Center, the GEO Group, Inc., or the Government of Canada. The authors wish to thank Leslie Helmus, Dan Montaldi, Donald R. Pake, Jr., and Pamela Yates for their helpful comments on earlier drafts of this manuscript.

² We concede that the Courts have yet to fully accept all recent changes to the Static-99 protocol. Indeed, the Judge in a recent New Hampshire Daubert hearing (New Hampshire v. Thomas Hurley, 07-E-0236) accepted the new scorings based on age, as well as the new recidivism estimates for “routine” offenders (thus, acknowledging that reoffense rates have decreased). She did not, however, admit evidence regarding use of normative samples other than routine, citing “(un)reliable principles and methods” in making assignments. Rectifying this situation will require addition empirical validation of the process, specifically regarding clear delineation of inclusion criteria for each normative group and the establishment of inter-rater reliability coefficients with respect to the assignment process.

References

- Abracen, J., & Looman, J. (2005). Developments in the assessment and treatment of sexual offenders: Looking backward with a view to the future. *Journal of Interpersonal Violence, 20*, 12-19.
- Andrews, D. A., & Bonta, J. (2000). *LSI-R: Level of Service Inventory—Revised*. Toronto, ON: Multi-Health Systems.
- Barbaree, H. E., Langton, C. M., & Blanchard, R. (2007). Predicting recidivism in sex offenders using the VRAG and SORAG: The contribution of age-at-release. *International Journal of Forensic Mental Health, 6*, 29-46.
- Campbell, T. W., & DeClue, G. (2010). Maximizing predictive accuracy in sexually violent predator evaluations. *Open Access Journal of Forensic Psychology, 2*, 148-232.
- Curtiss, P. R., & Warren, P. W. (1973). *The dynamics of life skills coaching*. Prince Albert, SK: Saskatchewan NewStart Inc. for the Training Research and Development Station Department of Manpower and Immigration.
- Federal Bureau of Investigation (2007). *Crime in the United States, 2007*. Available from <http://www.fbi.gov/ucr/cius2007/index.html>. (Accessed September 6, 2010)
- Finkelhor, D., & Jones, L. (2006). Why have child maltreatment and child victimization declined? *Journal of Social Issues, 62*, 685-716.
- Freund, K. (1963). A laboratory method for diagnosing predominance of homo- or hetero-erotic interest in the male. *Behaviour Research & Therapy, 1*, 85-93.
- Haaven, J. L., Little, R., & Petre-Miller, D. (1990). *Treating intellectual disabled sex offenders: A model residential program*. Orwell, VT: Safer Society.
- Hanson, R. K. (2006). Does Static-99 predict recidivism among older sexual offenders? *Sexual Abuse: A Journal of Research and Treatment, 18*, 343-355.
- Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). *A meta-analysis of the effectiveness of treatment for sexual offenders: Risk, need, and responsivity*. Research Report No. 2009-01. Ottawa: Corrections Research, Public Safety Canada.
- Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348-362.

- Hanson, R. K., & Harris, A. J. R. (2001). A structured approach to evaluating change among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 13, 105-122.
- Hanson, R. K., Harris, A. J. R., Scott, T. L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project*. [User Report 2007-05] Ottawa, ON: Public Safety Canada.
- Hanson, R. K., & Helmus, L. (2009, October). *Methods for combining historical and psychological risk factors: An example using Static-2002 and STABLE-2007*. Presentation at the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, Texas.
- Hanson, R. K., & Howard, P. D. (2010). Individual confidence intervals do not inform decision-makers about the accuracy of risk assessment evaluations. *Law and Human Behavior*, 34, 275-281.
- Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, 73, 1154-1163.
- 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, 1-21.
- Hanson, R. K., Phenix, A., & Helmus, L. (2009, September). *Static-99(R) and Static-2002(R): How to Interpret and Report in Light of Recent Research*. Paper presented at the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX. Available from www.static99.org. (Accessed September 6, 2010)
- Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sexual offenders*. [User Report 1999-02] Ottawa, ON: Department of the Solicitor General of Canada.
- Hanson, R. K., & Thornton, D. (2008, October). *Recommendations for interpreting multiple norms for the STATIC-99*. Presentation at the 27th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Atlanta, Georgia.
- Harkins, L., Thornton, D., & Beech, A. (2009, October). *The use of dynamic risk domains assessed using psychometric measures to revise relative risk assessment using RM 2000 and Static 2002*. Presentation at the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, Texas.

- Hare, R. (2003). *The Psychopathy Checklist—Revised* (2nd Edition). Toronto, ON: Multi-Health.
- Hare, R. D., & Neumann, C. S. (2010). The role of antisociality in the psychopathy construct: Comment on Skeem and Cooke (2010). *Psychological Assessment*, 22(2), 446-454.
- Harris, A. J. R., Helmus, L., Hanson, R. K., & Thornton, D. (2008, October). *Are New Norms Needed for Static-99?* Presented at the 27th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Atlanta, Georgia.
- Helmus, L. (2009). *Re-norming Static-99 recidivism estimates: Exploring base rate variability across sex offender samples* (Master's thesis). Available from ProQuest Dissertations and Theses database. (UMI No. MR58443).
- Helmus, L., Thornton, D., & Hanson, R. K. (2009, October). *Should Static-99 recidivism estimates be adjusted based on age at release? A multi-sample exploration.* Paper presented at the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX
- Levenson, J. S., & D'Amora, D.A. (2007). Social policies designed to prevent sexual violence: The emperor's new clothes? *Criminal Justice Policy Review*, 18, 168-199.
- Marlatt, G. A. (1998). *Harm reduction: Pragmatic strategies for managing high-risk behaviors*. New York: Guilford.
- Marshall, W. L. (2005). Clinical and research limitations in the use of phallometric testing with sexual offenders. *Sexual Offender Treatment*, 1, 14-41.
- Minnesota Department of Corrections (2007). *Sex offender recidivism in Minnesota: April 2007*. St. Paul, MN: Author.
- Monahan, J. (1981). *The clinical prediction of violent behavior*. Originally published by National Institute of Mental Health. Reprinted as *Predicting violent behavior: An assessment of clinical techniques* by Sage Publications, also 1981.
- Nafekh, M., & Motiuk, L. (2002). *The Statistical Information on Recidivism - Revised 1 (SIR-R1) Scale: A psychometric evaluation*. Research Report R-126. Ottawa, ON: Correctional Service of Canada.
- Nuffield, J. (1982). *Parole decision-making in Canada: Research towards decision guidelines*. Ottawa, ON: Communication Division.

- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk*. 2nd Edition. Washington, DC: American Psychological Association.
- Pake, D. R. (2010). There's no such thing as success in treatment. *ATSA Forum*, 22 (Summer).
- Phenix, A., Helmus, L., & Hanson, R. K. (2009). *Static-99R Evaluators' Handbook*. Available at www.static-99.org. (Accessed September 6, 2010).
- Skeem, J. L., & Cooke, D. J. (2010a). Is criminal behavior a central component of psychopathy? Conceptual directions for resolving the debate. *Psychological Assessment*, 22, 433-445.
- Skeem, J. L., & Cooke, D. J. (2010b). One measure does not a construct make: Directions toward reinvigorating psychopathy research - Reply to Hare and Neumann (2010). *Psychological Assessment*, 22, 455-459.
- Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research and Treatment*, 14, 139-153.
- Thornton, D., & Knight, R. (2009, October). *Using SRA need domains based on structured judgment to revise relative risk assessments based on Static-2002 and Risk Matrix 2000*. Presentation at the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, Texas.
- Thornton, D., Hanson, R. K., & Helmus, L. (2009). *Moving beyond the standard model for actuarial assessment for sexual offenders*. Available at www.ccoso.org/newsletter/movingbeyondstandard.doc. (Accessed September 6, 2010)
- Wilson, R. J., Looman, J., Abracen, J., & Pake, D. R. (2010, October). *Comparing sexual offenders at the Regional Treatment Centre and the Florida Civil Commitment Center*. Paper to be presented at the 29th Annual Conference of the Association for the Treatment of Sexual Abusers, Phoenix, AZ.
- Wilson, R. J., Picheca, J. E., & Prinzo, M. (2007). Evaluating the effectiveness of professionally-facilitated volunteerism in the community-based management of high risk sexual offenders: PART TWO—A comparison of recidivism rates. *Howard Journal of Criminal Justice*, 46, 327-337.
- Wilson, R. J., & Yates, P. M. (2009). Effective interventions and the Good Lives Model: Maximizing treatment gains for sexual offenders. *Aggression & Violent Behavior*, 14, 157-161.

Yates, P. M., & Ward, T. (2008). Good lives, self-regulation, and risk management: An integrated model of sexual offender assessment and treatment. *Sexual Abuse in Australia and New Zealand: An Interdisciplinary Journal*, 1, 3-20.