Forensic Use of the Static-99R: Part 3. Choosing a Comparison Group

Authors: Gregory DeClue & Denis L. Zavodny

Gregory DeClue, Sarasota, Florida
gregdeclue@mailmt.com (Corresponding author)

Denis L. Zavodny, Assistant Director, Office of Forensic Services
Georgia Department of Behavioral Health and Developmental Disabilities

Abstract

Although the developers of the Static-99R describe the "Selected for High Risk/Needs" comparison group as including offenders "referred for services at forensic psychiatric facilities, such as offenders referred as Mentally Disordered Sex Offenders, Sexually Violent Predators," no persons involved in SVP proceedings are included in any of the 23 research samples that comprise the four Static-99R comparison[1] groups. We compiled sexual-recidivism base-rate data from recent USA samples at various stages of the civil-commitment process. Available data do not consistently support the use of the "Selected for High Risk/Needs" comparison group for persons involved in the civil-commitment process. For example, Florida's SVP releases (persons who have been released after being found by Florida's Sexually Violent Predator Program to meet criteria for civil commitment) had 10-year detected sexual recidivism rates that were no higher than those of randomly selected sex offenders. We also consider the broader issue of selecting a Static-99R comparison group in the absence of local norms. We conclude that forensic evaluators should not choose a non-representative comparison group on the basis of clinical considerations unless and until empirical research demonstrates that such choices increase the accuracy of risk assessments.

Keywords: sexually violent predator, sexual recidivism, Static-99R, civil commitment

---

This is the third in a series of articles focused on the forensic use of the Static-99R. See DeClue (2013), and DeClue and Campbell (2013).

Beginning with Washington's law that went into effect in 1990 (Lieb & Matson, 1998), at least twenty states have instituted procedures for civil commitment of sexually violent predators (SVPs). In each state, psychological or psychiatric evaluations are a necessary part of the civil-commitment process, and an assessment of risk for sexual reoffense is an essential part of those evaluations. As Janus and Meehl (1997) note, the stakes are high because the consequences of the predictions are so severe. Prediction errors can be costly in terms of unnecessary, prolonged deprivations of lib-

[1] Throughout this article, we use the term "comparison group" instead of "norm group" because these are samples of convenience.

DeClue, G., & Zavodny, D. L. (2013) Forensic use of the Static-99R: Part 3. Choosing a comparison group. *Open Access Journal of Forensic Psychology, 5,* 151-182.

erty or a failure to prevent future sexual violence (Harris, 2005; LaFond, 2005). Therefore, it is important to maximize the accuracy of sexual-recidivism predictions in SVP cases (Campbell & DeClue, 2010).

The most researched and most widely used instrument for sexual-recidivism risk assessment has been the Static-99 (Hanson & Thornton, 2000). The instrument was revised in late 2009, and the developers now recommend the Static-99R for all purposes. At the official website of the Static-99 and related instruments,[2] the developers of the Static-99R present four different comparison groups. A July 2012 Workbook (Phenix, Helmus, & Hanson) posted at the official website describes one of the groups, the "Selected as High Risk/Needs" group, as including offenders "referred for services at forensic psychiatric facilities, such as offenders referred as Mentally Disorder (sic) Sex Offenders, Sexually Violent Predators . . ." (p. 36). Reading that description, an evaluator might choose the "Selected as High Risk/Needs" group as the appropriate comparison group for anyone involved in SVP proceedings. Our narrow purpose in this paper is to consider whether that choice is empirically supported. Our broad purpose is to consider how evaluators should choose a comparison group when using the Static-99R in forensic cases.

Background

RRASOR

Hanson and Bussière (1998) conducted a meta-analysis of factors associated with sexual re-offending. Using factors included in that research, Hanson (1997) developed a brief actuarial risk scale, the Rapid Risk Assessment for Sexual Recidivism (RRASOR). Using that tool with data sets available in the late 1990s, Hanson concluded, "The level of predictive accuracy found in this study suggest (sic) that it is possible to identify a large group of relatively low risk offenders whose chances of recidivism are less than 15% over ten years, as well as identifying a small group of sexual offenders whose chances of long-term recidivism are greater than 50%" (Hanson, 2007, p. 17).

Static-99

The Static-99 was developed by combining the RRASOR with another instrument, the SACJ-Min (Hanson & Thornton, 2000). "The Static-99 utilizes only static (unchangeable) factors that have been seen in the literature to correlate with sexual reconviction in adult males. The estimates of sexual and violent recidivism produced by the Static-99 can be thought of as a baseline of risk for violent and sexual reconviction" (Harris, Phenix, Hanson, & Thornton, 2003, p. 3). In other words, "Actuarial instruments can be seen as attempts to identify the base rate of re-offending for a specifiable sub-group of sex offenders" (DeClue, 2002, p. 76). Hanson and Thornton (2000) found that the 10-item Static-99 was more accurate than the 4-item RRASOR, but not by much: "The incremental improvement of Static-99, however, was relatively small" (p. 129). The developers of the Static-99 subsequently recognized that recidivism rates in later data sets were significantly lower than in the original samples and, in October 2008, recom-

[2] www.static99.org

mended the use of multiple comparison groups (Hanson & Thornton, 2008; Harris, Helmus, Hanson, & Thornton, 2008; Helmus, Hanson, & Thornton, 2009).

Static-99R

The Static-99 was revised in October 2009. "The developers of Static-99 recommend that the revised version of the scale (Static-99R) replace Static-99 in all contexts where it is used" (Phenix, et al., July 2012, p. 30). Static-99R has the same items as the original Static-99, except that the age item was revised to be more consistent with research findings regarding declines in sexual recidivism with increased age (Helmus, Thornton, Hanson, and Babchishin, 2012). For a second time, new comparison groups were provided after it was again recognized that sexual recidivism rates were much lower than in years past.

Static-99R Comparison Groups

Local Norms

To effectively use the Static-99R as an actuarial instrument, local norms are recommended (Helmus, Hanson, Thornton, Babchishin, & Harris, 2012; Phenix, et al., July 2012). Boccaccini, Murrie, Caperton, and Hawes (2009) discuss potential advantages of using local sexual-recidivism data for an actuarial instrument: "For example, when evaluating offenders released from a Texas prison, it would be preferable to consider recidivism in a large sample of Texas offenders, rather than to draw inferences from the large Static-99 dataset that includes relatively few U.S. samples, and no data specific to Texas prison inmates" (p. 284). Boccaccini and colleagues' (2009) study can be used as a model for collecting and reporting local norms for an actuarial instrument such as the Static-99R.

In most SVP states, local norms are not available. Therefore, in those states, evaluators refer to a non-local comparison group to help interpret the meaning of a particular score on the Static-99R. Four such groups are presented at the Static-99 website, recommended for use by evaluators, and are briefly described at this point.

A Group Representing the Full Population of Convicted Sex Offenders

In presenting the Static-99R, the developers provide a recidivism table for what they consider to be "samples representing the full population of all [convicted sex] offenders." In materials at www.static99.org, the developers refer to this comparison group as "Routine Correctional," "RC," or "Routine Samples." Here, we refer to the same group as FULLPOP, to keep the focus on the fact that they are considered to represent "the full population of all convicted sex offenders."[3]

[3] In our experience in SVP cases, we have heard experts testify that they did not use the "routine" comparison group because they did not consider the respondent to be "a routine sex offender."

A brief description of those eight groups follows[4]:

- sex offenders released from the Arizona (USA) Department of Corrections and subject to registration and notification (Bartosh, Garby, Lewis, & Gray, 2003)
- nearly all sex offenders receiving a federal sentence (two or more years) in Quebec, Canada, between 1995 and 2000 (Bigras, 2007)
- all male federal offenders serving a sentence for a sexual offense in British Columbia, Canada, whose sentence ended between January 1990 and May 1994 (Boer, 2003)
- all contact sex offenders on probation in two boroughs in South East London during a study period (Craissati, Bierer, & South, 2008)
- sex offenders released from prison in Austria (Eher, Rettenberger, Schilling, & Pfafflin, 2009)
- sex offenders in North Dakota who were either incarcerated or on probation (Epperson, 2003)
- offenders on community supervision in Canada during 2001-2005 (Hanson, Harris, Scott, & Helmus, 2007)
- sex offenders released from prison in Sweden (Långström, 2004)

Non-Representative Groups

At www.static99.org, the developers also present recidivism tables for some additional groups of samples that were not considered to be representative of the full population of all convicted sex offenders. The developers refer to these additional groups as "Non-routine" or "Pre-Selected." Here we refer to them as NONREP, to keep the focus on the fact that they are *not* considered to be representative of the full population of all convicted sex offenders. The Static-99R developers organized these non-representative samples into three groups: "Selected as Needing Treatment," "Selected as High Risk/Needs," and "Non-Routine." It is important to note that these are post-hoc groupings of disparate samples of convenience.

Selected as Needing Treatment. For this group, the Static-99R developers chose to combine six research samples, which included sex offenders released from

- a prison-based program in New Zealand (Allan, Grace, Rutherford, & Hudson, 2007)
- a maximum-security psychiatric facility in Quebec, Canada (Brouillette-Alarie & Proulx, 2008)
- community-based and prison-based treatment programs in the United Kingdom (Harkins & Beech, 2007)
- a prison-based sex-offender treatment program in Washington State, USA (Johansen, 2007)
- an outpatient treatment program at the University of Minnesota, USA (Swinburne Romine, Dwyer, Mathiowetz, & Thomas, 2008)

[4] These descriptions, and those that follow, are extracted from Phenix and colleagues (2012).

- a sex-offender treatment program in a minimum-security correctional facility in British Columbia, Canada (Ternowski, 2004)

Conceptually, the "Selected as Needing Treatment" group differed from the FULLPOP group in that they were all released from a sex-offender treatment program. It is not the case that none of the offenders in the FULLPOP group had been in a sex-offender treatment group; some of them had been and some had not.

Empirically, the detected sexual recidivism rate of the offenders grouped into the "Selected as Needing Treatment" group was higher (9.1% at five years) than that of those in the FULLPOP group (6.0% at five years).[5]

Selected as High Risk/Needs. For this group, the Static-99R developers chose to combine six research samples, which included

- sex offenders who received a pre-trial forensic psychiatric evaluation in Denmark, typically due to some indication of high risk (Bengtson, 2008)
- offenders flagged as potential Dangerous Offenders, subject to indeterminate sentence, by the Canadian National Flagging System (Bonta & Yassine, 2005)
- Canadian federal sex offenders who were denied (routine) early release due to a significant risk of committing a serious offense prior to the time when their sentence would expire (Haag, 2005)
- offenders who were assessed or treated at a Massachusetts Treatment Center for sexually dangerous persons between 1959 and 1984 (Knight & Thornton, 2007)
- sex offenders treated at a maximum-security forensic mental-health facility in Saskatchewan, Canada (Nicolaichuk, 2001)
- another set of Canadian sex offenders who were denied (routine) early release due to a significant risk of committing a serious offense prior to the time when their sentence would expire (Wilson, Cortoni, & Vermani, 2007; Wilson, Picheca, & Prinso, 2007)

Conceptually, the "Selected as High-Risk/Needs" group differed from the FULLPOP group in that they were all released from a specialized facility or program for high-risk persons. It is not the case that none of the offenders in the FULLPOP group had been psychiatrically evaluated or had maxed out their prison sentences; some of them had and some had not.

Empirically, the detected sexual recidivism rate of the offenders grouped into the "Selected as High-Risk/Needs" group was higher (21.0% at five years) than that of those in the FULLPOP group (6.0% at five years).[6]

[5] Detailed recidivism tables Static-99R (October 2009). Available at www.static99.org
[6] Detailed recidivism tables Static-99R (October 2009). Available at www.static99.org

Non-Routine. For this group, the Static-99R developers chose to combine fifteen research samples, which included the six samples in the "Selected as Needing Treatment" group, the six samples in the "Selected as High-Risk/Needs" group, plus three additional samples that consisted of

- imprisoned sex offenders in Canada's National Sexual Offender Treatment Program (Cortoni & Nunes, 2007)
- offenders who committed a sexual homicide, from German federal criminal records (Hill, Haberman, Klusmann, Berner, & Briken, 2008)
- sex offenders treated by North Dakota's Department of Human Services (Saum, 2007)

Conceptually, the "Non-Routine" group differed from the FULLPOP group in that they were all released from a specialized facility or program for high-risk persons, or were released from a sex-offender treatment program, or had committed a sexual homicide. It is not the case that none of the offenders in the FULLPOP group had been in sex-offender treatment, had been psychiatrically evaluated, or had maxed out their prison sentences; some of them had and some had not.

Empirically, the detected sexual recidivism rate of the offenders grouped into the "Non-Routine" group was higher (14.8% at five years) than that of those in the FULLPOP group (6.0% at five years).[7]

The Practical Impact of the Choice of a Comparison Group

Several criticisms of the multiple-comparison-group process have previously been raised. There is noteworthy score variation within each Static-99R reference group (Abbott, 2013), which potentially compromises the utility of the new comparison groups (Sreenivason, Weinberger, Frances, & Cusworth-Walker, 2010). In addition, the developers have been criticized for providing insufficient details about the samples (Sreenivason et al., 2010) and for criterion contamination due to overlapping characteristics between groups (Abbott, 2011). The insertion of a subjective step into an "actuarial" risk-assessment tool can be seen as providing "legitimizing cover" (Prentky, Janus, Barbaree, Schwartz, & Kafka, 2006, p. 361) or a "veneer of 'quantification'" (Sreenivasan, et al., 2010, p. 405) to clinical judgment.

As we turn to the practical impact of the choice of a comparison group, we begin by noting that most states do not have local norms available for the Static-99R. In our experience with SVP cases, we have never seen the use of local norms by evaluating experts.

Across the four Static-99R comparison groups, not only is the base rate of sexual offending different for each group, but also the risk level for each score of the Static-99R is different across groups (Helmus, et al., 2009). Those differences are so substantial that the choice of comparison group is often the one that has the most impact on the

[7] Detailed recidivism tables Static-99R (October 2009). Available at www.static99.org

outcome of the "actuarial" risk assessment. For example, an evaluator who uses the recommended tables (Phenix, et al., July 2012) would find a predicted 5-year sexual recidivism rate of 8.7% if she used the FULLPOP (routine) comparison group. If, instead, she chose the "Selected as High Risk/Needs" comparison group, she would find a predicted 5-year sexual recidivism rate of 20.1% and a predicted 10-year sexual recidivism rate of 29.6%. How high would a person's Static-99R score have to be to reach a predicted sexual recidivism rate of 29.6% using the FULLPOP comparison group? Because the developers of the Static-99R have not provided predicted 10-year sexual recidivism rates for the FULLPOP comparison group, evaluators only have the 5-year data available. For the FULLPOP group, the highest Static-99R with an associated predicted recidivism rate is 9. The predicted sexual recidivism rate is 29.5, the highest for any score in the FULLPOP group. So, the practical impact of choosing the "Selected as High Risk/Needs" group is as substantial as a change in Static-99R score from 4 to 9.[8]

Static-99R Developers' Recommendations Regarding Comparison Groups

To effectively use the Static-99R as an actuarial instrument, local norms are recommended (Helmus, et al., 2012; Phenix et al., July 2012), but, as mentioned above, in most SVP cases, local norms are not available.

The developers of the Static-99R state that, when local norms are not available, "The routine [FULLPOP] sample will usually reflect the most appropriate recidivism rates…" (Phenix, et al., July 2012; p. 32). However, there is little to guide an evaluator who is tempted to consider using one of the non-representative comparison groups: "Deciding which set of norms is appropriate can be challenging" (Thornton, 2011, p. 1) and "The best method of determining which sample type is the most appropriate match to an individual is not fully known" (Phenix, et al., July 2012, p. 35). The added task of selecting a comparison group would insert clinical judgment into the interpretation of the Static-99R (Storey, Watt, Jackson, & Hart, 2012), thereby decreasing the reliability of the risk assessment (Abbott, 2011; Abbott, 2013; Wollert, 2010). At present, there are no empirical studies showing that an evaluator can produce a more accurate risk assessment by using a non-representative comparison group, rather than using the comparison group that is considered to be representative of the full population of convicted sex offenders (FULLPOP).

The developers of the Static-99R mention two primary issues that evaluators might consider regarding choice of comparison group: (a) clinical consideration of "psychologically meaningful risk factors" (Phenix et al., July 2012, p. 35) and (b) consideration of "the features of the three samples" (Phenix et al., July 2012, p. 35). We will discuss each of those issues after an extended focus on a misleading factoid that has guided testimony in some SVP cases in which we have participated.

[8] To consider the impact of the choice of comparison group for other Static-99R scores, see pages 5 to 8 of Phenix, et al. (July 2012).

A Misleading Factoid

The July 2012 Workbook (Phenix, et. al.) posted at the official website describes one of the groups, the "Selected as High-Risk/Needs" group, as including offenders "referred for services at forensic psychiatric facilities, such as offenders referred as Mentally Disorder (sic) Sex Offenders, Sexually Violent Predators . . ." (p. 36). Reading that description, an evaluator might choose the "Selected as High-Risk/Needs" group as the appropriate comparison group for anyone involved in SVP proceedings. In fact, though, no persons involved in any stage of any SVP proceedings were included in any of the six research samples that comprise the "Selected as High-Risk/Needs" group (R. K. Hanson, personal communication, April 12, 2012).

An Educated Guess?

Although there are no persons who were involved in SVP proceedings in the "Selected as High-Risk/Needs" group, or in any of the four comparison groups at the Static-99R website, the developers' mentioning of SVP in a discussion of a "Selected as High-Risk/Needs" group could be treated as some type of guess. That is, the developers of the Static-99R are apparently guessing that people referred for SVP proceedings would have risk levels comparable to those of the sex offenders in the six samples that the developers decided to group together with the label, "Selected as High-Risk/Needs." However, such an approach is tautological.

There are at least two reasons why evaluators should be especially cautious about choosing a comparison group on the basis of a guess, rather than empirical findings. First, there is a growing body of research that demonstrates that forensic evaluators may be at risk of bias, intentional or unintentional, related to allegiance effects (Murrie, Boccaccini, Guarnera, & Rufino, (2013); Murrie, Boccaccini, Turner, Meeks, Woods, & Tussy, 2009). Generally, the more subjective the decision faced by evaluators, the more likely the evaluators' ratings or test scores will align themselves with those of the side (plaintiff or respondent) who hired them, consistent with "myside" (Stanovich, West, & Toplak, 2013) or confirmation (Nickerson, 1998) bias.

Second, the persons in the "Selected as High-Risk/Needs" group are not drawn from general population of a state prison, which, in our experience, is the source of most SVP referrals. One of those samples, Knight and Thornton (2007), contributed the highest number of cases to the "Selected as High-Risk/Needs" group. That study followed offenders who were assessed or treated at a specialized treatment center for sexually dangerous persons between 1959 and 1984, so the 5-year window for their sexual recidivism data would have been in the 1960s, 1970s, and 1980s -- time periods when the observed sexual-recidivism rates were much higher than they are for sex offenders recently released from confinement in the USA. In fact, between 1990 and 2004, there was a 49 percent decline in observed sexual abuse (Finkelhor & Jones, 2006). Therefore, although the developers of the Static-99R may have reasons for their guess that people in the SVP process would have sexual recidivism rates comparable to those in the "Selected as High-Risk/Needs" group, there are reasons for guessing otherwise.

Obviously, it would be better to make such decisions on the basis of empirical research rather than guessing.

Is There Any Evidence to Support or Contradict That Guess?

Whether people involved in SVP proceedings sexually recidivate at a level comparable to the group labeled "Selected as High-Risk/Needs" is an empirical question. Because none of the 23 Static-99R studies at www.static99.org includes persons who had been involved in SVP proceedings, we sought existing data sets that would be relevant to SVP evaluators and decision makers. We compiled sexual-recidivism base-rate data from recent USA samples at various stages of the civil-commitment process.

In order to understand the data, we need to consider the detected sexual recidivism rates of (a) the persons in the 23 samples that were used in the development of the Static-99R (presented in the Appendix), (b) the four groups devised by the developers of the Static-99R (groups of those 23 samples) (presented in the Appendix), and (c) recent USA samples of sex offenders released from confinement (presented below).

Sexual Recidivism Rates from Recent USA Samples

In our quest to find data relevant to cases in the SVP process, we requested data sets through the Sex Offender Civil Commitment Programs Network,[9] and by contacting forensic directors of states with civil-commitment programs.[10] We also considered data sets that were provided to us by researchers and practitioners who became aware that we were collecting such data sets. Finally, we submitted a public records request to Florida's Sexually Violent Predator Program (SVPP).

We received several recent USA studies that report relevant base rates. All except the Florida SVPP study have been previously published and/or made publicly available via the Internet. Not all have been published in peer-reviewed journals. The studies include recent findings from California, Connecticut, Florida, Minnesota, Texas, South Carolina, New Jersey, and Washington, with sexual recidivism base rates varying from 0.8% to 25.2%. In Table 1, results are organized to maximize the usefulness for SVP risk assessments.[11] Such base rates provide crucial information for sexual-recidivism risk assessments in SVP cases.

For seven states, Connecticut, Texas, Washington, Minnesota, Florida, New Jersey, and South Carolina, there are data regarding the detected sexual recidivism of sex offenders released from prison after serving a sentence for a sex offense. There are 5-year detected sexual recidivism rates for three of those states, ranging from 2% to 7.0% (see Table 1).

[9] www.soccpn.org/home.html
[10] www.nasmhpd.org/About/rosters/Forensic92612.pdf
[11] A previous version of this table was presented as a poster at the 2012 convention of the American Psychological Association in Orlando, Florida.

Table 1: Sexual Recidivism Base Rates in Some Recent US Samples

Status	State/ Jurisdiction	Observed Sexual Recidivism Rate	Length of Follow-up
Released from prison after serving a sentence for a sex offense	CT[12]	27/746 (3.6%) charged; 20/746 (2.7%) convicted	5 years
	TX[13]	2.0% (mandatory supervision); 5.5% (no mandatory supervision); 3.4% (all sex offenders)	5 years
	WA[14]	2.7%	5 years
	MN[15]	12% arrested; 10% convicted	average of 8.4 years
	MN[16]	7.0% 12.9%	5 years 10 years
	FL[17]	5.2% 13.7%*	5 years 10 years
	NJ[18]	3.5% 8.3%	5 years 10 years
	SC[19]	4.1% 7.0%	5 years 10 years
Screened by SVP/SDP program and referred for a face-to-face evaluation; after evaluation, evaluator recommends civil commitment	TX[20]	0.8% (mandatory supervision) 7.5% (no mandatory supervision)	2.25 - 7.5 years (M = 4.77, SD = 1.52)
	WA[21]	34/135 (25.2%)	6 years
Found to meet SVP criteria – or – probable cause had been found – or – two evaluators had found the person to meet SVP criteria; released without treatment	CA[22]	6/93 (6.5%)	4.7 years
Found by SVP Program to meet criteria for civil commitment, subsequently released[23]	FL[24]	6/155 (4%) 14/134 (10%) 30/251 (12%) 21/170 (12%)**	0-3 years 3-5 years 5-10 years 10-14 years
Civilly committed as SVP; subsequently judicially released without completing SVP treatment program	FL[25]	2/39 (5.1%) (any sex-related charge) 2/39 (5.1%) (felony charge for sexually motivated offense)	0-10+ years[26]
Civilly committed as SVP; subsequently judicially released after completing SVP treatment program[27]	FL[28]	5/61 (8.2%) (any sex-related charge) 2/61 (3.3%) (felony charge for sexually motivated offense)	0-10+ years

Notes Regarding Table 1

[12] Recidivism Among Sex Offenders in Connecticut, 2/15/12.
[13] Boccaccini, Murrie, Caperton, and Hawes (2009).
[14] Barnoski (2005).
[15] Sex Offender Recidivism in Minnesota, April 2007.
[16] Zgoba et al. (2012). Note that all the subjects in this study were randomly selected sex offenders who were released from prison.
[17] Zgoba et al. (2012).
[18] Zgoba et al. (2012).
[19] Zgoba et al. (2012).
[20] Boccaccini et al. (2009).
[21] Milloy (2007).
[23] Recidivism rates are for any sexually motivated offense, including charges or convictions.
[24] Montaldi, Lewis, & Heffron (2013).
[25] Montaldi et al. (2013).
[26] For the 100 total subjects in the last two rows of this table, 51 subjects were at risk for 0-3 years, 22 subjects were at risk for 3-5 years, 20 subjects were at risk for 5-10 years, and 7 subjects were at risk for 10+ years (Montaldi et al., 2013).
[27] Here, "released after completing SVP treatment" is defined as released while on Phase IV (follow-up) of the SVP treatment program.
[28] Montaldi et al. (2013).
* Randomly selected sex offenders, 10 years at risk.
** Sex offenders identified by Florida's SVPP as meeting criteria for civil commitment, 10-14 years at risk.

We found four data sets that are particularly relevant to the guess that people in the SVP process would have detected sexual recidivism rates similar to those in the "Selected as High Risk/Needs" group. Data from two states, Texas and Washington, included persons who had been screened by an SVP program and referred for a face-to-face evaluation; after the evaluation, the evaluator recommended civil commitment. Typically, these would be the cases that would go to a civil commitment trial. For the Texans, fewer than 1% of those who were released with mandatory supervision were detected to sexually re-offend within the 2.25- to 7.5-year follow-up period. For the Texans who were released without mandatory supervision, 7.5% were detected to sexually recidivate within the same follow-up period. Results were quite different for offenders released from Washington at a similar stage in the SVP process, with 25.2% being detected to sexually recidivate within a 6-year follow-up period (see Table 1, above).

Another data set particularly relevant to SVP evaluators and decision makers is a group from California who were released from confinement and followed for 4.7 years. This group consisted of persons who were found to meet SVP criteria, or probable cause for such had been found, or two evaluators had found the person to meet SVP criteria. None of them had been civilly committed and treated. As shown in Table 1, 6.5% of them were detected to sexually recidivate within 4.7 years.

Data presented in Table 1 regarding Floridians show that SVP releases (persons who have been released after being found by Florida's Sexually Violent Predator Program to meet criteria for civil commitment) had 10-year detected sexual recidivism rates that

were no higher than those of randomly selected sex offenders (See data points marked * and **).

Data were sparse regarding persons who have been civilly committed as sexually violent predators and subsequently released. As shown in the last two rows of Table 1, the only data set we found were for people who had been committed as sexually violent predators, and who were subsequently judicially released after completing an SVP treatment program at the Florida Civil Commitment Center. Persons released after completing SVP treatment had similar detected sexual-recidivism rates as those who were released without having completed treatment.

Conclusion Regarding the Use of the "Selected as High Risk/Needs" Comparison Group because the Person Has Been Referred for SVP Proceedings or Commitment

The Static-99R developers' guess is not consistently supported by available empirical research. The most relevant samples are the Texans and Washingtonians who were "screened by SVP/SDP program and referred for a face-to-face evaluation; after evaluation, evaluator recommends civil commitment," the Californians who were "found to meet SVP criteria – or – probable cause had been found – or – two evaluators had found the person to meet SVP criteria; released without treatment," and the Floridians who were released after a determination by Florida's SVP Program that they met criteria for civil commitment. The Washingtonians recidivated at a rate (25.2% over 6 years) that is comparable to the "Selected as High Risk/Needs" group, but the Floridians, Texans, and Californians recidivated at rates much closer to those of the FULLPOP group. In fact, only about 1% of those Texans on mandatory supervision were found to sexually recidivate within an average of 4.77 years, so even using the FULLPOP group would over-predict their risk of detected sexual recidivism.

The short answer to our narrow question is that the Static-99R developers' guess regarding persons involved in SVP proceedings is not consistently supported by the available empirical research. Unless and until further research shows otherwise, evaluators should not consider that the results of an SVP screening and referral process (e.g., opinions that the person meets criteria for civil commitment) warrant use of the "Selected as High Risk/Needs" comparison group.

The data in Table 1 address one aspect of the choice of a non-representative comparison group in the absence of local norms. These data fail to consistently support the guess that persons involved in the SVP civil-commitment process should be compared to the "Selected as High Risk/Needs" comparison group.

We turn now to two recommendations regarding choice of a non-representative comparison group, and then we will close with a more general discussion of the choice of a non-representative comparison group.

Recommendations to Select a Comparison Group Based on Clinical Considerations

Documents at www.static99.org provide at least two recommendations regarding choice of a non-representative comparison group: (a) clinical consideration of "psychologically meaningful risk factors"[29] and (b) consideration of "the features of the three samples."[30]

As mentioned above, the first recommendation regarding choice of a comparison group is to use local norms. When local norms are not available, the FULLPOP (routine) comparison group is typically the next best choice. "Given that these norms are not often available, *the routine sample will usually reflect the most appropriate recidivism rates as they are representative of typical sex offenders in correctional systems"* (Phenix et al., July 2012, p. 32, emphasis added). We consider the above description to be reasonable and sound, given what is known and what is not known about sexual recidivism, as long as the sample has been systematically collected and is sufficiently large so that it is representative of the group.

Now consider the guidelines in the same manual regarding the non-representative groups (p. 35): "Determining the appropriate sample type reference group should be based on consideration of psychologically meaningful risk factors (e.g., a dynamic risk assessment scale). . . . In the absence of information about external psychologically meaningful risk factors, evaluators can also consider the features of the three samples."

Clinical consideration of "psychologically meaningful risk factors." Even if evaluators could follow such a recommendation reliably,[31] there has been no research showing that this clinical process leads to more accurate (valid) risk assessments than if the evaluator simply used the FULLPOP group (which is considered to represent the full population of all sex offenders). In fact, there have been five studies showing that, for sexual-recidivism risk assessments, when people use their judgment to arrive at a risk estimate different from the standard rate, that decreases the accuracy of the risk assessment (Gore, 2007; Hanson, 2007; Storey, et al., 2012; Vrana, Sroga, & Guzzo, 2008; Wormith, Hogg, & Guzzo, 2012; see also Campbell & DeClue, 2010; DeClue, 2013; Hanson & Morton-Bourgon, 2009).

Regarding the first three studies, all were prospective. Hanson and Morton-Bourgon (2009, p. 7) summarized the results of the earlier three studies as follows: "In two studies, the raters were probation officers (Hanson, 2007; Vrana et al., 2008), and in the other study, the raters were either psychologists or correctional staff (Gore, 2007). For all three measures, for all types of raters, and for all outcomes, the adjusted scores showed lower predictive accuracy than did the unadjusted actuarial scores."[32]

---

[29] July 2012 Evaluators Workbook at www.static99.org, p. 35.
[30] July 2012 Evaluators Workbook at www.static99.org, p. 35.
[31] Anecdotally, one of us recently testified in an SVP case in which four evaluators testified that they used the Static-99R. Each of the four evaluators considered the correct Static-99R score to be 5, but various evaluators selected the FULLPOP (routine) group, preselected-for-treatment group, or high-risk/high-needs group. The associated positive predictive values (PPVs) varied from .16 to .37.
[32] See Campbell and DeClue (2010) for discussion of these three studies.

Two additional studies, both published in 2012, provide further data. Storey et al. (2012) looked at 100 cases in which evaluators scored the Static-99.[33] In 30 of the cases, the evaluators made a clinical decision to report "risk estimates that differ from the standard rates, due to the presence of case-specific moderating factors" (p. 5), referred to as a "clinical override." Storey et al. (p. 8) found "the clinical override scores were less predictive of sexual recidivism than the scores without overrides [and] the ratings with overrides predicted recidivism in the wrong direction – that is, clinical overrides of increased risk were actually associated with lower recidivism rates and vice versa."

Wormith, Hogg, and Guzzo (2012) examined the predictive validity of the Level of Service/Case Management Inventory (LS/CMI) on a sample of sexual offenders extracted from a large cohort of offenders in Canada. "The study revealed that allowing assessors to override the numerically derived risk level with their professional judgment, a practice that increased risk level much more often than it decreased it, reduced the predictive validity of the scale and did so particularly for sex offenders by increasing risk excessively" (p. 1511). "When the validity of the LS/CMI was examined for sex offenders on whom the override provision was applied, the predictive relationship was eliminated" (p. 1530). "In our view, these findings illustrate the potential shortcomings of using SPJ [structured professional judgment] to augment a statistically based risk/need assessment scheme" (p. 1531). "The current findings do not auger well for the professional override having overall incremental predictive validity, at least as employed with the current guidelines in the LS/CMI. Rather, they partially support Abbott's (2011) position that professional judgment should not be used to increase actuarially based assessments of sexual offenders' risk of recidivism" (p. 1533).

In theory, it is possible that a standardized procedure could be developed whereby evaluators would use a dynamic risk-assessment tool in addition to a static-factor tool such as the Static-99R. Next, it could be tested whether carefully trained evaluators in a controlled study, using that combination of tools, arrive at more accurate predictions than evaluators who relied solely on the static-factor actuarial tool. A third step would be field studies to address the practical impact of using the combination procedure in actual cases. Even if well-trained evaluators could use the procedure effectively under controlled conditions, it would be important to explore whether allegiance or other social-psychological factors decrease the accuracy of risk assessments in forensic cases (see, e.g., Boccaccini et al., 2009).

At present, there is no research showing that incremental validity is added by using clinical judgment regarding "external psychologically meaningful risk factors" to augment or facilitate a statistically based risk-assessment scheme.[34] In practice, SVP evaluators, who would choose to use "external psychologically meaningful risk factors" to select a

[33] All of the clinicians had received extensive training in use the Static-99 and the Sex Offender Need Assessment Rating (SONAR) prior to making their ratings.

[34] Some initial steps have been tried. For example, Sachsenmaier, Thornton, and Olson (2011), attempted to use the Structured Risk Assessment (SRA), but they found unsatisfactorily low inter-rater reliability.

comparison group with a higher sexual-recidivism base rate, would be using "external psychologically meaningful risk factors" to arrive at a higher estimated risk for sexual reoffense. Although this recommendation comes via the intermediate step of selecting a comparison group, the practical result is to use professional judgment to arrive at a risk estimate other than the standard, actuarial-based rate. There is no specific, empirical evidence to support this recommendation. Further, it is contraindicated by five recent, empirical studies that show that inserting professional judgment regarding external factors leads to a decrease in the accuracy of actuarial-based risk assessment.

Consideration of "the features of the [comparison groups]." This section focuses on the following quote from the July 2012 Workbook at www.static99.org: "In the absence of information about external psychologically meaningful risk factors, evaluators can also consider the features of the three samples[35]" (p. 35).

This is similar, but not identical to, standard recommendations for considering whether an actuarial tool is appropriate for a particular risk assessment. For example, the persons in the development sample for the Static-99 were all adult males. Therefore, if the evaluation task is to assess the risk of future sex offending and the person to be evaluated is a female, then the Static-99 would not be an appropriate tool to use. Nor would it be appropriate for assessing a person who committed his most recent sex offense at age 14. The same considerations hold true for the Static-99R.

With the Static-99R, the developers provide a comparison group that they identify as "the closest available to a hypothetical average of all [adult male] sex offenders." We consider it to be obvious that, in conducting a risk assessment that employs the Static-99R, it would always be useful to include a comparison to the FULLPOP group. Whenever it is useful to score the Static-99R, it would be important to consider the overall recidivism rate of sex offenders with that person's score. The question for a forensic evaluator, then, is when to refer to a non-representative comparison group *in addition to* the FULLPOP group.

Our recommendation is that evaluators can refer to a non-representative comparison group, *in addition to* the FULLPOP group, when there is clear evidence that the objective characteristics of the *sample* from which the evaluee is drawn are similar to the objective characteristics of one or more of the non-representative samples. For example, suppose an evaluee was referred for SVP consideration after maxing out his sentence (losing all of his gain time due to misbehavior) in a maximum-security facility. Those objective facts are similar to the factual descriptions of the "Selected as High Risk/Needs" group on pages 22-23 of Helmus et al. (2012). Therefore, in this hypothetical case, we would recommend that an evaluator mention the sexual recidivism rates of the "Selected as High Risk/Needs" group *in addition to* the sexual recidivism rates of the FULLPOP group.

[35] Actually, there were 23 samples, which were sorted into four groups on a post-hoc basis. The post-hoc nature of the sorting is particularly evident in that one sample, Saum (2007), was excluded from the "Selected as Needing Treatment" comparison group "because it was a statistical outlier" (Helmus et al., 2012, p. 24).

General Recommendations. In providing three non-representative comparison groups known to have higher sexual-recidivism rates than the FULLPOP group, the developers of the Static-99R create the opportunity for evaluators to choose a higher-risk comparison group and, consequently, arrive at a higher risk estimate for the person being evaluated. As mentioned previously, these risk estimates are often much higher – even two or three times higher – than the estimated risk that results from a comparison to the full population of sex offenders. For example, for a Static-99R score of 5, the 5-year estimated risk for sexual reoffense for the full population of sex offenders is 11.4 (p. 5 of the July 2012 Workbook).[36] For the same score using the "Selected as High Risk/Needs" comparison group, the 5-year estimated risk is 25.2, and the 10-year estimated risk is 35.5 (p. 8 of the July 2012 Workbook).

As mentioned above, the developers of the Static-99R caution, "Using any of the non-routine or preselected norms, however, requires justification" (Phenix et al., 2012, p. 36). Currently, we do not believe that the choice of a non-representative group can be justified on the basis of clinical considerations, for two reasons. First, there is zero empirical research showing increased accuracy by switching to a non-representative group. Second, extant research consistently shows that clinical adjustments from the standard actuarial rate decrease accuracy of sexual-recidivism risk assessments.

Overall Conclusions for the Present

It is not hard to imagine a study that would provide a field test of the www.static99.org guidelines regarding choice of a comparison group. A study could be modeled after Storey et al. (2012). A sample of risk assessments would be identified wherein evaluators had scored the Static-99R and chosen whichever comparison group they deemed appropriate, using the guidelines available at www.static99.org. Sexual recidivism data would be collected for sex offenders released from confinement. For those cases in which the evaluator had chosen a non-representative comparison group, researchers would explore whether actual sexual recidivism rates more closely approximate those of the non-representative group chosen by the evaluator, versus the rate from the FULLPOP group.

While we await the results of such a study, and other research, our recommendations are very similar to those of Storey et al. (2012), who recommend, "Additional and more detailed guidelines regarding the appropriate use of overrides should be tested empirically and provided to clinicians. Alternatively, clinicians should be discouraged from

[36] The 5-year rate is the only one provided for the FULLPOP comparison group. For the other comparison groups, both a 5-year rate and a 10-year rate are provided. Hanson (Hanson & Phenix, 2013) stated that the reason was that they did not have enough 10-year data for the FULLPOP group. In fact, though, they had 10-year sexual recidivism data for 750 persons in the samples comprising the FULLPOP group, which they did not report, compared to 722 persons in the High-Risk group, which they did report. It remains unclear why the Static-99R developers have chosen to report 10-year sexual-recidivism data for the High-Risk comparison group, but not for the FULLPOP comparison group. See the Appendix.

overriding Static-99 scores under any circumstances." Our recommendations for forensic evaluations are as follow:

Regarding the use of Static-99R in forensic cases:

1. *Guidelines for choosing a non-representative comparison group on the basis of clinical considerations should be tested empirically before being used by a forensic evaluator. Unless and until such choices are found to increase the accuracy of risk assessments, forensic evaluators should use local norms (if available) or the FULLPOP comparison group (considered roughly representative of all adjudicated sex offenders).*

2. *Forensic evaluators who use the Static-99R in the absence of local norms should always refer to the FULLPOP comparison group. Evaluators could refer to a non-representative comparison group, in addition to the FULLPOP group, when there is clear evidence that the objective characteristics of the sample from which the evaluee is drawn are similar to the objective characteristics of one or more of the non-representative samples.*

Future Directions

At a recent annual conference of the Association for the Treatment of Sexual Abusers, Hanson and Phenix (2013) discussed their current recommendations regarding the use of the Static-99R, and they mentioned some developments currently underway. During the presentation, Hanson acknowledged a key point that we have presented in this article: There has been no research showing that evaluators choosing a comparison group leads to more accurate risk predictions, compared to using the comparison group based on the full population of convicted sex offenders.

Phenix described a recent history of successive sets of recommendations for choosing a comparison group. The initially recommended procedure "was a matching process based on level of pre-selection for an offender." Problems in that recommended procedure included a lack of fit between referral source and characteristics of the developmental samples, inadequate inter-rater reliability in choosing the norms, and potential for bias affecting forensic evaluators due to vagueness in the procedure.

The second recommended procedure was to shift from considering a pre-selection process to assessing the observable levels of "external risk factors" or "dynamic needs." Phenix currently recommends that evaluators use a "structured, quantitative method" to measure the person's "external risk factors" or "dynamic needs," and then use those results to select a Static-99R comparison group. This is likely to result in greater inter-rater reliability among evaluators. However, Phenix acknowledged that there is no research demonstrating more-accurate risk predictions for evaluators who use a structured instrument to guide selection of a Static-99R comparison group. Indeed, throughout the four years that evaluators have been encouraged to select a Static-99R

comparison group via one procedure or another, there has never been evidence that doing so increases predictive accuracy.

Hanson and Phenix (2013) mentioned that, based on recently analyzed data, they anticipate abandoning two of the three non-representative Static-99R comparison groups: the "Selected as Needing Treatment" and "Non-Routine" comparison groups. They anticipate that in the near future they will recommend the use of two, rather than four, comparison groups. One comparison group will be randomly selected sex offenders, or the near-equivalent thereof. The other will be a high-risk comparison group. If/when the Static-99R developers describe a new procedure for choosing between those two comparison groups, forensic evaluators will need to carefully consider whether the new procedure has been shown to lead to more accurate risk predictions than relying on the comparison group that represents the full population of convicted sex offenders.

Received May 7, 2013; revision submitted August 16, 2013; second revision submitted October 31, 2013.

Appendix: Sexual Recidivism Rates from Static-99R Development Samples and Comparison Groups

Sexual Recidivism Rates of the 23 Samples

Table 2 presents 5- and 10-year sexual recidivism rates for the 23 studies used in the development of the Static-99R. This table was developed from data in Helmus' (2009, p. 39) Table 8. This is Helmus' master's thesis, which sparked the development of the Static-99R. Her thesis is available on the Static-99 website.

Table 2: Detected Sexual Recidivism Rates for the Static-99R Samples
[Adapted from Helmus' Table 8 (2009, p. 39)]

Sample	Comparison Group	N		Detected to Have Sexually Recidivated			
		5-Year	10-Year	5-Year		10-Year	
				N	%	N	%
Cortoni & Nunes (2007)	NONREP	17	X	0	0.0	X	X
Hanson et al. (2007)	FULLPOP	31	X	0	0.0	X	X
Eher et al. (2008)	FULLPOP	151	X	3	2.0	X	X
Boer (2003)	FULLPOP	299	295	11	3.7	23	7.8
Långström (2004)	FULLPOP	1278	353	69	5.4	26	7.4
Johansen (2007)	NONREP, Treatment	272	62	16	5.9	8	12.9
Ternowski (2004)	NONREP, Treatment	247	X	16	6.5	X	X
Craissati et al. (2008)	FULLPOP	200	66	15	7.5	6	9.1
Swinburne Romine et al. (2008)	NONREP, Treatment	570	543	48	8.4	61	11.2
Bigras (2007)	FULLPOP	207	X	19	9.2	X	X
Harkins & Beech (2007)	NONREP, Treatment	198	129	19	9.6	21	16.3
Epperson (2003)	FULLPOP	151	36	16	10.6	8	22.2
Hill et al. (2008)	NONREP	73	54	8	11.0	10	18.5
Allan et al. (2007)	NONREP, Treatment	299	25	35	11.7	5	20.0
Wilson et al. (2007a & b)	NONREP, High Risk	104	16	12	11.5	1	6.3
Bartosh et al. (2007)	FULLPOP	90	X	12	13.3	X	X
Brouillette-Alarie & Proulx (2008)	NONREP, Treatment	199	110	29	14.6	23	20.9
Bonta & Yessine (2005)	NONREP, High Risk	101	3	19	18.8	0	0.0
Bengtson (2008)	NONREP, High Risk	310	291	61	19.7	83	28.5
Haag (2005)	NONREP, High Risk	198	X	39	19.7	X	X
Nicholaichuk (2001)	NONREP, High Risk	168	59	38	22.6	15	25.4
Knight & Thornton (2007)	NONREP, High Risk	433	353	107	24.7	106	30.0
Saum (2007)	NONREP	175	X	55	31.4	X	X

Notes Regarding Tables 2-7

FULLPOP = Samples included in the comparison group considered to be representative of the full population of convicted sex offenders.

Treatment = Samples included in the comparison group labeled “Selected as Needing Treatment.”

High Risk = Samples included in the comparison group labeled “Selected as High-Risk/Needs.”

NONREP = Samples not included in the FULLPOP comparison group. The NONREP comparison group includes the six samples labeled “Selected as Needing Treatment” and the six samples labeled “High-Risk/Needs” plus three additional samples.

N = number of persons
X = No such cases were included in the sample.

Although the Static-99R was developed from the 23 studies listed in Table 2, the developers excluded some cases from the samples due to coding inconsistencies or missing items (Helmus, et al., 2012). Above, in Table 2, we present detected sexual recidivism rates from the more complete data set.

Because it may sometimes be useful for evaluators to refer only to data that made its way into the Static-99R data sets, we present Table 3, below, based on the data reported in Helmus et al. (2012, p. 71), which appears to be the first peer-reviewed article introducing the Static-99R.

Table 3: Detected Sexual Recidivism Rates for the Static-99R Samples
[Adapted from Helmus et al. (2012, p. 71)]

Sample	Comparison Group	N		Detected to Have Sexually Recidivated			
		5-Year	10-Year	5-Year		10-Year	
				N	%	N	%
Cortoni & Nunes (2007)	NONREP	17	X	0	0.0	X	X
Hanson et al. (2007)	FULLPOP	31	X	0	0.0	X	X
Eher et al. (2008)	FULLPOP	151	X	3	2.0	X	X
Boer (2003)	FULLPOP	299	295	11	3.7	23	7.8
Långström (2004)	FULLPOP	1278	353	69	5.4	26	7.4
Johansen (2007)	NONREP, Treatment	272	62	16	5.9	8	12.9
Ternowski (2004)	NONREP, Treatment	247	X	16	6.5	X	X
Craissati et al. (2008)	FULLPOP	200	66	15	7.5	6	9.1
Swinburne Romine et al. (2008)	NONREP, Treatment	569	542	48	8.4	61	11.3
Bigras (2007)	FULLPOP	207	X	19	9.2	X	X
Harkins & Beech (2007)	NONREP, Treatment	197	127	19	9.6	21	16.5
Epperson (2003)	FULLPOP	150	36	16	10.7	8	22.2
Hill et al. (2008)	NONREP	73	54	8	11.0	10	18.5
Allan et al. (2007)	NONREP, Treatment	298	25	35	11.7	5	20.0
Wilson et al. (2007a & b)	NONREP, High Risk	103	16	12	11.7	1	6.3
Bartosh et al. (2007)	FULLPOP	90	X	12	13.3	X	X
Brouillette-Alarie & Proulx (2008)	NONREP, Treatment	199	110	29	14.6	23	20.9
Bonta & Yessine (2005)	NONREP, High Risk	81	3	14	17.3	0	0.0
Bengtson (2008)	NONREP, High Risk	310	291	61	19.7	83	28.5
Haag (2005)	NONREP, High Risk	198	X	39	19.7	X	X
Nicholaichuk (2001)	NONREP, High Risk	168	59	38	22.6	15	25.4
Knight & Thornton (2007)	NONREP, High Risk	433	353	107	24.7	106	30.0
Saum (2007)	NONREP	169	X	50	29.6	X	X

Sexual Recidivism Rates of the Four Comparison Groups

Next, we present detected sexual recidivism rates for the four post-hoc comparison groups devised by the Static-99R developers (Phenix, et al., 2012). The more complete data set is presented as Table 4, below. Table 4 is derived from the data in Table 2.

Table 4: Detected Sexual Recidivism Rates for the Static-99R Comparison Groups Using Helmus (2009) Data

Comparison Group	Number of Cases		Detected to Have Sexually Recidivated			
	5-Year	10-Year	5-Year		10-Year	
			N	%	N	%
FULLPOP	2407	750	145	6.0	63	8.4
Treatment	1785	869	163	9.1	118	13.6
High Risk	1314	722	276	21.0	205	28.4
NONREP	3364	1645	502	14.9	333	20.2

Table 5, below, includes only the data that made its way into the Static-99R as described in Helmus et al. (2012). Table 5 is derived from the data in Table 3 (which were drawn from Helmus et al., 2012).

Table 5: Detected Sexual Recidivism Rates for the Static-99R Comparison Groups Using Helmus et al.'s (2012) Data

Comparison Group	Number of Cases		Detected to Have Sexually Recidivated			
	5-Year	10-Year	5-Year		10-Year	
			N	%	N	%
FULLPOP	2406	750	145	6.0	63	8.4
Treatment	1782	866	163	9.1	118	13.6
High Risk	1293	722	271	21.0	205	28.4
NONREP	3334	1642	492	14.8	333	20.3

The "Norms" tab at www.static99.org provides a link to a document dated November 2009: Detailed recidivism tables Static-99R.[37] The values from that document are presented in Table 5, below. As can be seen, the values in Table 6 do not exactly match those of either Table 4 or Table 5.

[37] http://www.static99.org/pdfdocs/detailed_recid_tables_static99r_2009-11-15.pdf

Table 6: Detected Sexual Recidivism Rates for the Static-99R Comparison Groups Using Totals from Detailed recidivism tables Static-99R at www.static-99.org

Comparison Group	Number of Cases		Detected to Have Sexually Recidivated			
	5-Year	10-Year	5-Year		10-Year	
			N	%	N	%
FULLPOP	2406	X	145	6.0	X	X
Treatment	1777	866	167	9.4[a]	118	13.6
High Risk	1313	703	276	21.0	204	29.0[b]
NONREP	3354	1626	497	14.8	332	20.4

[a]Reported as 9.1%
[b]Reported as 29.1%

The developers of the Static-99R chose not to report 10-year data for the FULLPOP group in the Detailed recidivism tables Static-99R (see Tables 6 and 7). Hanson (Hanson & Phenix, 2013) stated that the reason they did not report 10-year data for the FULLPOP group is that they did not have enough data. However, examination of Tables 4 and 5 reveals that the developers actually had 10-year recidivism data for 750 persons in the samples comprising the FULLPOP group, compared to 722 persons in the samples comprising the High-Risk group. Yet the Static-99R developers chose to report 10-year sexual recidivism rates for the High-Risk group, but not for the FULLPOP group.

As noted above, the Detailed Recidivism Tables at www.static99.org appear to have two calculation errors regarding the sexual recidivism rates. Table 6, above, provides detected sexual recidivism percentages that are consistent with the totals reported in the final rows of each of the Detailed Recidivism Tables. For example, for the 5-year sexual recidivism data for the "Selected as Needing Treatment" group, the "totals" row reports that 167 of 1,777 persons were detected to have sexually recidivated, which is 9.4%, not 9.1% as reported in the Detailed Recidivism Tables.

Another problem with the Detailed Recidivism Tables is that the columns do not always add up to the value reported as the total. For example, for the 5-year sexual recidivism data for the "Selected as Needing Treatment" group, the data points in the Detailed Recidivism Table sum to 167 of 1,782 persons detected to have sexually recidivated, not 167 of 1,777. Table 7, below, presents data consistent with the data points in the Detailed Recidivism Tables.

Table 7: Detected Sexual Recidivism Rates for the Static-99R Comparison Groups Using Data Points from Detailed recidivism tables Static-99R at www.static-99.org

Comparison Group	Number of Cases		Detected to Have Sexually Recidivated			
	5-Year	10-Year	5-Year		10-Year	
			N	%	N	%
FULLPOP	2406	X	145	6.0	X	X
Treatment	1782	866	163	9.1	118	13.6
High Risk	1313	703	276	21.0	204	29.0
NONREP	3354	1626	497	14.8	332	20.4

Summary Regarding Sexual Recidivism Rates from Static-99R Development Samples and Comparison Groups

The differences between Tables 2 and 3 are minor and would not be expected to have much practical impact. The same can be said for differences among Tables 4 through 7. Nevertheless, these discrepancies highlight the value of having an official manual for any tool that is to be used in forensic cases, so that there can be general agreement among those who use the tool. Any errors, once recognized, can be acknowledged and officially corrected. For forensic evaluators using the Static-99R, the evaluator should know which of the above data sets he or she is using, and be prepared to provide a rationale for using that one rather than one of the alternatives. The decision not to report 10-year sexual-recidivism rates for the FULLPOP comparison groups remains unexplained.

References

Abbott, B. R. (2011). Throwing the baby out with the bath water: Is it time for clinical judgment to supplement actuarial risk assessment? *Journal of the American Academy of Psychiatry and Law, 39*, 222-230.

Abbott, B. R. (2013). The utility of assessing "external risk factors" when selecting Static-99R reference groups. *Open Access Journal of Forensic Psychology, 5,* 89-118.

Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric assessment of dynamic risk factors for child molesters. *Sexual Abuse: A Journal of Research and Treatment, 19,* 347-367.

Barnoski, R. (2005). *Sex offender sentencing in Washington State: Recidivism rates.* Olympia: Washington State Institute for Public Policy, Document No. 05-08-1203.

Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy & Comparative Criminology, 47*, 422-438.

Bengtson, S. (2008). Is newer better? A cross-validation of the Static-2002 and the Risk Matrix 2000 in a Danish sample of sexual offenders. *Psychology, Crime & Law, 14,* 85-106.

Bigras, J. (2007). La prédiction de la récidive chez les délinquants sexuels [Prediction of recidivism among sex offenders]. *Dissertations Abstracts International, 68* (09). (UMI No. NR30941).

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, 2009, 15,* 278-314.

Boer, A. (2003). *Evaluating the Static-99 and Static-2002 risk scales using Canadian sexual offenders*. Unpublished master's thesis, University of Leicester, Leicester, United Kingdom.

Bonta, J., & Yessine, A. K. (2005). [Recidivism data for 124 released sexual offenders from the offenders identified in *The National Flagging System: Identifying and responding to high-risk, violent offenders* (User Report 2005-04). Ottawa: Public Safety and Emergency Preparedness Canada]. Unpublished raw data.

Brouillette-Alarie, S., & Proulx, J. (2008, October). *Predictive and convergent validity of phallometric assessment in relation to sexual recidivism risk.* Poster presented at the annual conference for the Association for the Treatment of Sexual Abusers, Atlanta, GA.

Campbell, T. W., & DeClue, G. (2010). Flying blind with naked factors: Problems and pitfalls in adjusted-actuarial sex-offender risk assessment. *Open Access Journal of Forensic Psychology, 2,* 75-101.

Cortoni, F., & Nunes, K. L. (2007). *Assessing the effectiveness of the National Sexual Offender Program* (Research Report No. R-183). Unpublished report, Correctional Service of Canada.

Craissati, J., Bierer, K., & South, R. (2008). *What do sex offenders really get up to? Risk prediction, community failure and "sexually risky behaviours" in a nine year follow up study.* Unpublished manuscript.

DeClue, G. (2002). Avoiding garbage in sex offender re-offense risk prediction: A case study. *Journal of Threat Assessment, 2,* 73-92.

DeClue, G. (2013). Years of predicting dangerously. *Open Access Journal of Forensic Psychology, 5,* 16-28.

DeClue, G., & Campbell, T. W. (2010). Still maximizing accuracy in sexually violent predator evaluations. *Open Access Journal of Forensic Psychology, 2,* 322-336.

DeClue, G., & Campbell, T. W. (2013). Calibration performance indicators for the Static-99R: 2013 update. *Open Access Journal of Forensic Psychology, 5,* 81-88.

Eher, R., Rettenberger, M., Schilling, F., & Pfafflin, F. (2008). Failure of Static-99 and SORAG to predict relevant reoffense categories in relevant sexual offender subtypes: A prospective study. *Sexual Offender Treatment, 8*(1), 1-20.

Epperson, D. L. (2003). *Validation of the MnSOST-R, Static-99, and RRASOR with North Dakota prison and probation samples*. Unpublished Technical Assistance Report, North Dakota Division of Parole and Probation.

Finkelhor, D., & Jones, L. (2006). Why have child maltreatment and child victimization declined? *Journal of Social Issues, 62*, 685-716.

Gore, K. S. (2007). Adjusted actuarial assessment of sex offenders: The impact of clinical overrides on predictive accuracy. *Dissertation Abstracts International, 68*(07), 4824B. (UMI No. 3274898).

Haag, A. M. (2005). [recidivism data from 198 offenders detained until their warrant expiry date. From: Do psychological interventions impact on actuarial measures: An analysis of the predictive validity of the Static-99 and Static-2002 on a re-conviction measure of sexual recidivism*. Dissertations Abstracts International, 66* (08), 4531B. (UMI No. NR05662)]. Unpublished raw data.

Hanson, R. K. (1997). The development of a brief actuarial risk scale for sexual offense recidivism. (User Report 97-04). Ottawa: Department of the Solicitor General of Canada.

Hanson, R. K. (March 2007). *How should risk assessments for sexual offenders be conducted?* Paper presented at the Fourth Annual Forensic Psychiatry Conference, Victoria, British Columbia, Canada. Cited in R. K. Hanson & K. Morton-Bourgon. (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., & 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., Scott, T.-L., & Helmus, L. (2007). *Assessing the Risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Corrections Research User Report No. 2007-05). Ottawa, ON, Canada: Public Safety Canada.

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. (2013). Report writing for Static-99R and Static-2002R. Pre-conference seminar at the 32nd Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Chicago, IL, October 30, 2013.

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

Hanson, R. K., & Thornton, D. (2008, October). *Recommendations for Interpreting Multiple Norms for the Static-99*. Paper presented at the 27th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Atlanta, Georgia.

Harris, A. J. (2005). *Civil commitment of sexual predators: A study in policy implementation*. New York: LFB Scholarly Publishing.

Harris, A. J. R., Phenix, A., Hanson, R. K., & Thornton, D. (2003). Static-99 coding rules: Revised 2003. Ottawa, ON: Solicitor General Canada.

Harris, A. J., Helmus, L., Hanson, R. K., & Thornton, D. (2008, October). *Are new norms needed for the Static-99?* Paper 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 at www.static99.org

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

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.

Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse: Journal of Research and Treatment, 24,* 64-101.

Hill, A., Habermann, N., Klusmann, D., Berner, W., & Briken, P. (2008). Criminal recidivism in sexual homicide perpetrators. *International Journal of Offender Therapy and Comparative Criminology, 52,* 5-20.

Janus, E. S., & Meehl, P. E. (1997). Assessing the legal standard for prediction of dangerousness in sex offender commitment hearings. *Psychology, Public Policy, and Law, 3*(1)*,* 33-64.

Johansen, S. H. (2007). Accuracy of predictions of sexual offense recidivism: A comparison of actuarial and clinical methods. *Dissertations Abstracts International, 68* (03), B. (UMI No. 3255527).

Knight, R. A., & Thornton, D. (2007). *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders* (Document No. 217618). Submitted to the U.S. Department of Justice.

LaFond, J. Q. (2005). *Preventing sexual violence: How society should cope with sex offenders. Washington, D.C.: American Psychological Association.*

Långström, N. (2004). Accuracy of actuarial procedures for assessment of sexual offender recidivism risk may vary across ethnicity. *Sexual Abuse: A Journal of Research and Treatment, 16*, 107-120.

Lieb, R., & Matson, S. (1998). *Sexual predator commitment laws in the United States: 1998 update.* Olympia: Washington State Institute for Public Policy.

Milloy, C. (2007). Six-year follow-up of 135 released sex offenders recommended for commitment under Washington's sexually violent predator law, where no petition was filed. Olympia: Washington State Institute for Public Policy, Document No. 07-06-1101. Accessed July 28, 2012, from www.wsipp.wa.gov/pub.asp?docid=07-06-1101

Mondaldi, D. F., Lewis, S., & Heffron, J. (2013, August). Sexually Violent Predator Program recidivism study: Preliminary results. Presentation to a panel completing a program review requested by the Florida Department of Children and Families, Substance Abuse and Mental Health (SAMH) Program Office, Tallahassee, Florida.

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychological Science, 24,* 1889-1897.

Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussy, 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 Law, 15,* 19-53.

Nicholaichuk, T. (2001, November). *The comparison of two standardized risk assessment instruments in a sample of Canadian Aboriginal sexual offenders.* Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Antonio, TX.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175-220.

Phenix, A., Helmus, L., & Hanson, R. K. (July 2012). *Static-99R & Static-2002R Evaluators' Workbook*. Retrieved from www.static99.org

Prentky, R. A., Janus, E., Barbaree, H., Schwartz, B. K., & Kafka, M. P. (2006). Sexually violent predators in the courtroom. *Psychology, Public Policy, and Law, 12*, 357-393.

Recidivism Among Sex Offenders in Connecticut, State of Connecticut Office of Policy and Management, 2/15/12. Accessed July 28, 2012 from http://www.ct.gov/opm/cwp/view.asp?a=2976&q=383710

Sachsenmaier, S., Thornton, D., & Olson, G. (2011, November). *Structured Risk Assessment–Forensic Version: Score Distribution, Inter-rater Reliability, and Margin of Error in an SVP Population.* Paper presented at the 30th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Toronto.

Saum, S. (2007). A comparison of an actuarial risk prediction measure (Static-99) and a stable dynamic risk prediction measure (Stable-2000) in making risk predictions for a group of sexual offenders. *Dissertations Abstracts International, 68* (03), B. (UMI No. 3255539).

Sex Offender Recidivism in Minnesota, April 2007. Minnesota Department of Corrections. Accessed July 28, 2012, from http://www.doc.state.mn.us/documents/04-07SexOffenderReport-Recidivism.pdf

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*, 400-4006.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science, 22,* 259-264.

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

Swinburne Romine, R., Dwyer, S. M., Mathiowetz, C., & Thomas, M. (2008, October). *Thirty years of sex offender specific treatment: A follow-up Study*. Poster presented at the conference for the Association for the Treatment of Sexual Abusers, Atlanta, GA.

Ternowski, D. R. (2004). Sex offender treatment: An evaluation of the Stave Lake Correctional Centre Program. *Dissertations Abstracts International, 66* (06), 3428B. (UMI No. NR03201).

Thornton, D. (2011). *Notes for Using SRA-FV Level of Need Index*. Unpublished Manuscript.

Vrana, G. C., Sroga, M., & Guzzo, L. (2008). *Predictive validity of the LSI–OR among a sample of adult male sexual assaulters.* Unpublished manuscript, Nipissing University, North Bay, Ontario, Canada.

Wilson, R. J., Cortoni, F., & Vermani, M. (2007). *Circles of support and accountability: A national replication of outcome findings* (Report N., R-185). Ottawa, ON: Connectional Service of Canada.

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. *The Howard Journal, 46*, 327-337.

Wollert, R. (2010, March). *The use of probability mathematics in sexual violent predator (SVP) evaluations.* Workshop presented at the California Department *of Mental Health, Seaside.*

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.

Zgoba, K. M., Miner, M., Knight, R., Letourneau, E., Levenson, J., & Thornton, D. (2012). A multi-state recidivism study using Static-99R and Static-2002 risk scores and tier guidelines from the Adam Walsh Act. Research report submitted to the National Institute of Justice, Award Number 2008-MU-MU-000.