

Forensic Use of the Static-99R:

Part 4. Risk Communication

Gregory DeClue

Sarasota, Florida

Denis L. Zavodny

Georgia Department of Behavioral Health and Developmental Disabilities

Author Note

Correspondence concerning this article should be sent to gregdeclue@me.com

Abstract

The Static-99R is one of the most widely used tools for assessing sexual-recidivism risk. We draw on recent research and the Specialty Guidelines for Forensic Psychologists to provide succinct answers to six basic questions regarding the use of the Static-99R in forensic cases. We suggest that evaluators focus on detected sexual recidivism at the group level, using concepts from the field of public health and safety, expressed in everyday terms. In the absence of local norms, we recommend that evaluators refer to the comparison group that is considered to represent the full population of convicted sex offenders. We encourage evaluators to affirmatively state that there is no evidence that clinical adjustments or overrides enhance the accuracy of actuarial-based risk predictions. We advise evaluators to affirmatively state that there is no objective procedure that can allow an evaluator to reliably and validly predict that someone is more-likely-than-not to sexually recidivate if he is not confined. We present some cross-examination questions that forensic evaluators should be prepared to answer.

Keywords: Static-99R, forensic psychology, actuarial risk assessment, sex-offender risk assessment, professional guidelines

The likelihood of future sexual reoffending is a primary consideration in some forensic cases, including those involving civil commitment of sexually violent predators (SVPs; Jackson & Richards, 2008; Miller, Amenta, & Conroy, 2005; Rogers & Shuman, 2005; Schwartz, 1999). Actuarial tools have been readily adopted in risk assessments, including assessments of the risk for sexual re-offense (Richards, 2013; Tully, Chou, & Browne, 2013). Prior to its revision in 2009, the Static-99 (Hanson & Thornton, 2000) was considered to be the most widely used actuarial instrument for that purpose. The developers of the Static-99 now recommend use of the revised version, the Static-99R, for all purposes, and offer guidelines not only for scoring the instrument, but also for how the results could be interpreted (Phenix, Helmus, & Hanson, July 26, 2012). Although it has now been four years since the Static-99R was introduced, too little attention has been focused on the use of the Static-99R in forensic cases.

This is the fourth in a series of articles focused on the forensic use of the Static-99R. See DeClue (2013), DeClue and Campbell (2013), and DeClue and Zavodny (2013). In this article, we consider how evaluators who use the Static-99R in forensic cases can communicate the results in a way that is meaningful, useful, and consistent with research. After presenting some brief background information about the Static-99R, we consider pertinent sections of the Specialty Guidelines for Forensic Psychologists (American Psychological Association, 2013). We then address the following six basic questions:

1. Should I report risk for sexual recidivism, or risk for *detected* sexual recidivism?
2. Should I report individual, or group, risk?
3. Which group risk should I report?
4. How can risk be clearly and accurately conveyed to fact finders?

5. Do clinical adjustments or overrides enhance the accuracy of sexual-recidivism risk predictions?
6. Is there any objective procedure that can allow me to reliably and validly predict that someone is more-likely-than-not to sexually recidivate if not confined?

Background

RRASOR

Hanson and Bussière (1998) conducted a meta-analysis of factors associated with sexual reoffending. Applying variables identified 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). It is important to recognize that (a) when the modern era of sexual-recidivism risk assessment began, researchers believed that it was possible to objectively identify a small group of sex offenders whose long-term likelihood of sexual recidivism exceeded 50%, but (b) current research shows that this cannot be done in an objective manner. This fact will become evident as we track some developments over the ensuing years.

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 reoffending 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 as “The incremental improvement of Static-99, however, was relatively small” (p. 129).

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 26, 2012, p. 30). The 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 decline in sexual recidivism with increased age (Helmus, Thornton, Hanson, and Babchishin, 2011). It was also recognized that sexual recidivism rates are much lower than in years past. In presenting the Static-99R, the developers provided a recidivism table for what they consider to be “samples representing the full population of all [convicted sex] offenders.”¹ In materials at the official website of the Static-99 and related instruments, www.static99.org, the developers refer to this comparison group as “Routine Correctional,” “RC,” or “Routine Samples.” Here, as in prior papers in this series, 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.

¹ “This group consisted of eight samples of sex offenders from Canada, the United States, England, Austria, and Sweden. These samples were relatively random (i.e., unselected) samples from a correctional system (as opposed to samples from specific institutions or subject to specific measures). In other words, they can be considered roughly representative of all adjudicated sex offenders. Some offenders in these samples would have been subsequently screened for treatment or other special measures (e.g., psychiatric admission or exceptional measures related to dangerousness), but these samples represent the full population of all offenders prior to any preselection processes. The recidivism norms for the unselected samples are the closest available to a hypothetical average of all sex offenders” (Phenix, et al., July 26, 2012, p. 32).

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. For the FULLPOP comparison group, only 5-year sexual recidivism rates are available. Tables 1 and 2 were developed from data available at www.static99.org.

Insert Table 1 about here.

Noting that the raw data show a lack of smooth progression from lower scores to higher scores,² the developers of the Static-99R used a statistical procedure³ to develop estimates of what they expect the data would show if the instrument were used with larger samples.⁴ Table 2 shows the observed sexual recidivism rates, along with estimates of what the recidivism rates might be if more subjects were included. Note that the confidence intervals are for the subjects in the FULLPOP group, not for a risk estimate for any single person (see Hart & Cooke, 2013). Those confidence intervals were calculated by the developers of the Static-99R and are taken directly from their website.

Insert Table 2 about here.

Specialty Guidelines for Forensic Psychologists

In this section, we identify pertinent sections of the 2013 Specialty Guidelines (American Psychological Association, 2013). In the following section, we draw upon these guidelines as we consider six basic questions regarding the forensic use of the Static-99R.

Guideline 2.05: Knowledge of the Scientific Foundation for Opinions and Testimony

² For example, the detected recidivism rate for a score of 5 is higher than it is for a score of 6.

³ Logistic regression.

⁴ See page 5 of 1/9/2012 “Static-99R & Static-2002R Evaluators’ Workbook” at <http://www.static99.org/>.

This guideline consists of the following: “Forensic practitioners seek to provide opinions and testimony that are sufficiently based upon adequate scientific foundation, and reliable and valid principles and methods that have been applied appropriately to the facts of the case. When providing opinions and testimony that are based on novel or emerging principles and methods, forensic practitioners seek to make known the status and limitations of these principles and methods” (p. 9). Forensic practitioners should always keep this guideline in mind. It is particularly pertinent to questions 3, 5, and 6, below. Forensic practitioners need to understand which aspects of Static-99R interpretation have been empirically validated, and which have not. This knowledge should be clearly presented in our risk communications.

Guideline 10.02: Selection and Use of Assessment Procedures

This guideline includes, “Forensic practitioners use assessment instruments whose validity and reliability have been established for use with members of the population assessed. When such validity and reliability have not been established, forensic practitioners consider and describe the strengths and limitations of their findings” (p. 15). This guideline, too, is particularly pertinent to questions 3, 5, and 6, below.

Guideline 11.01: Accuracy, Fairness, and Avoidance of Deception

This guideline includes, “Forensic practitioners make reasonable efforts to ensure that the products of their services, as well as their own public statements and professional reports and testimony, are communicated in ways that promote understanding and avoid deception. When in their role as expert to the court or other tribunals, the role of forensic practitioners is to facilitate understanding of the evidence in dispute. Consistent with legal and ethical requirements, forensic practitioners do not distort or withhold relevant evidence or opinion in reports or testimony” (p. 16). For example, forensic practitioners who testify about Static-99R results must

not withhold the fact that adjustments or overrides decrease the accuracy of sex-offender risk assessments. See question 5, below.

Guideline 11.02: Differentiating Observations, Inferences, and Conclusions

This guideline includes, “In their communications, forensic practitioners strive to distinguish observations, inferences, and conclusions” (p. 16). For example, when explaining the use of the Static-99R for a person in a legal case, a forensic evaluator should distinguish between (a) coding rules, (b) suggestions regarding interpretation, and (c) any intermediate steps. For actuarial instruments, it is important to adhere precisely to the coding rules, so that information about an evaluatee is scored the same way it would have been scored if he had been in the development sample. Suggestions for interpretation should be followed *when they are empirically supported*. The same holds true for intermediate steps, such as suggestions for choosing a comparison group. See question 3, below.

Guideline 11.04: Comprehension and Accurate Presentation of Opinions in Reports and Testimony

This guideline includes, “Consistent with relevant law and rules of evidence, when providing professional reports and other sworn statements or testimony, forensic practitioners strive to offer a complete statement of all relevant opinions that they formed within the scope of their work on the case [including] the basis and reasoning underlying the opinions” (p. 17). The core of an actuarial risk assessment can be conducted with a great deal of objectivity. However, the same forensic practitioner who presents objective findings from an actuarial risk assessment may also present opinions that are essentially subjective in nature. To avoid misrepresentation and misunderstanding, it is essential that forensic practitioners clearly identify when the opinions they state are based on subjective reasoning.

Six Basic Questions Regarding Forensic Use of the Static-99R

1. Should I report risk for sexual recidivism, or risk for detected sexual recidivism?

Specialty Guidelines 11.01 and 11.04 are pertinent. Recidivism research does not report how many people sexually reoffended. It reports how many people were detected to sexually reoffend, within a particular time frame. We do not know how many people in a research sample sexually recidivated. We know how many people were detected to sexually recidivate, using a particular measure of detected recidivism (e.g., arrest or conviction for a new sex offence). Therefore, evaluators should communicate risk for *detected* sexual recidivism.⁵

2. Should I report individual risk or group, risk?

Specialty Guidelines 10.02, 11.01, and 11.04 are pertinent. Two recent articles provide persuasive data and arguments, convincing us that evaluators should report group, not individual, risk predictions. We encourage readers to carefully consider the data and analysis in the original articles. Here, we summarize some of their conclusions.

Using data from a follow-up study of 90 adult male sex offenders, Hart and Cooke (2013) developed a new actuarial risk-assessment instrument. Their focus was not so much on predictions of how likely a person is to sexually recidivate, but on how precise the predictions can be. They illustrate that an actuarial risk-assessment tool can moderately and significantly predict sexual recidivism in the aggregate, but “group probability estimates had substantial margins of error and individual probability estimates had very large margins of error” (p. 81).

Hart and Cooke persuasively show that the lack of precision is not a limitation in one sample or one tool, but is endemic to attempts to make such predictions about individuals. “We conclude that, without major advances in our understanding of the causes of violence, ARAIs

⁵ Although there have been some proposals for estimating actual recidivism from detected recidivism, these proposals have been seriously flawed (Wollert, 2006; Wollert & Cramer, 2011).

[actuarial risk assessment instruments] cannot be used to estimate the specific probability or absolute likelihood of future violence with any reasonable degree of precision or certainty” (Hart & Cooke, 2013, p. 81). Further, the imprecision does not only occur with actuarial tools: “Of course, clinicians cannot use SPJ [structured professional judgment] guidelines to make individual risk estimates in the form of specific probabilities or absolute likelihood with any more precision than they can with ARAIs. Clinicians should consider whether it is best to give up altogether on the idea of calculating probability estimates of risk for future violence” (Hart & Cooke, 2013, p. 98).

Another recent study took a decidedly different route to arrive at a very similar conclusion. Fazel, Singh, Doll, & Grann (2012) conducted a systematic review and meta-analysis of tools commonly used to assess the risk of violence, including sexual violence. The entire study included 73 samples, with 24,847 people from 13 countries. Risk-assessment tools included both actuarial instruments focusing on historical risk factors, and tools based on clinical judgment. For both types of tools, the accuracy of the instruments was not sufficient to support a precise prediction that an individual is likely to sexually recidivate: “In some criminal justice systems, expert testimony commonly use[s] scores from these instruments in a simplistic way to estimate an individual’s risk of serious repeat offending. However, our review suggests that risk assessment tools in their current form can only be used to roughly classify individuals at the group level, and not to safely determine criminal prognosis in an individual case” (Fazel et al., 2012, p. 5).

The upshot is that evaluators should be careful to convey the fact that our risk assessments are imprecise. This is especially important because the expression of risk in numerical form may suggest that it can be assessed at a misleadingly high degree of precision

(Sreenivasan, Weinberger, Francis, & Cusworth-Walker, 2010). Although our tools and procedures allow for better-than-chance classification of groups of people, the level of certainty regarding an individual prediction is quite low (Scurich & John, 2012a). This appears to be consistent with the following recommendation from Melton, Petrila, Poythress, and Slobogin (2007, p. 320): “If the examiner uses actuarial risk assessment tools, . . . a precise probability estimate can be provided. . . . Of course, such estimates, if presented to the factfinder, should be carefully explained for what they are: recidivism rates for people in a particular population with specific characteristics that the examinee shares.” When it comes to individual predictions about which individuals will sexually recidivate and which individuals will not, we do not know.

3. Which group risk should I report?

Specialty Guidelines 2.05, 10.02, 11.01, 11.02, and 11.04 are pertinent. We concur with the developers of the Static-99R that, when possible, evaluators should use local norms to interpret the meaning of a person’s Static-99R score. Generally, when local norms are not available, evaluators should use the FULLPOP comparison group, the one considered to represent the full range of convicted sex offenders.⁶ We recommend that evaluators should not choose a specialized comparison group on the basis of clinical considerations unless and until empirical research demonstrates that such choices increase the accuracy of risk assessments.

At the time of this writing (late 2013), there is no body of research demonstrating any reliable and valid way of using any of the NONREP comparison groups in a forensic case. When an evaluator scores a Static-99R for a particular person and there are no local norms available, the evaluator can use the FULLPOP comparison group to compare that person’s score to a sample representing the full population of convicted sex offenders. That is a straightforward

⁶ At the time of this writing (Summer 2013), in materials at www.static99.org, this comparison group is referred to as “Routine Correctional,” “RC,” or “Routine Samples.”

and potentially useful process. Although the developers of the Static-99R have provided alternate comparison groups, there has been no empirical evidence that evaluators who use a specialized comparison group arrive at more accurate risk predictions, compared to using the FULLPOP comparison group (Abbott, 2013). If an evaluator uses one of the NONREP comparison groups, we know that the risk prediction will be higher (greater likelihood to sexually reoffend) compared to a prediction based on the FULLPOP group, but there is no evidence that the risk prediction will be more accurate. Currently, the NONREP comparison groups at www.static99.org are contra-indicated for forensic evaluations because their use involves inherent bias with no empirical justification. For a detailed analysis, see DeClue and Zavodny (2013).

4. How can risk be clearly and accurately conveyed to fact finders?

Specialty Guidelines 11.01 and 11.04 are pertinent. In addressing the prior three questions, we recognized that, in the absence of local norms, evaluators should report data regarding detected sexual recidivism within a group representing the full range of convicted sex offenders. The framing of risk probabilities can dramatically influence the perception of risk (Gigerenzer, 2002; Scurich & John, 2011; Slovic, Monahan, & MacGregor, 2000), and, consequently, it may be beneficial to note both the estimated likelihood of reoffending coupled with its converse, the estimated likelihood of not reoffending (Scurich & John, 2011).

Recognizing the strengths and limitations of each metric, in the *Evaluators' Workbook* (Phenix, et al., July 26, 2012), the developers of the Static-99R recommend the use of absolute recidivism rates (Helmus, Hanson, Thornton, Babchishin, & Harris, 2012), percentiles (Hanson, Lloyd, Helmus, & Thornton, 2012), and relative risk ratios (Hanson, Babchishin, Helmus, & Thornton, 2013) for quantifying and communicating risk. They suggest that relative risk,

providing an index of likelihood of reoffending compared to other offenders, is a more fundamental attribute than percentiles or absolute risk (Babchishin, Hanson, & Helmus, 2012; Hanson, et al., 2013). However, relative risk ratios, percentiles, and labels such as “high,” “medium,” and “low” (compared to other sex offenders) may be misleading and likely result in misinterpretation or usurpation of decision-making by fact finders, typically leading to an overestimation of risk (Cook & Michie, 2014; Gigerenzer, 2002; Gigerenzer & Gray, 2011; Krauss, McCabe, & Lieberman, 2012; Krauss & Scurich, 2014; Hanson, et al., 2013; Scurich & John, 2012b). Therefore, in any proceedings in which estimated absolute risk is of interest (viz. sexually violent predator trials), we do not recommend that evaluators report relative risk ratios, percentiles, or labels such as “high,” “medium,” or “low.”

Schopp (2001) suggested that the role of the expert should be limited to description and explanation, and the role of the trier of fact is to make a value judgment after weighing competing interests. We recommend that evaluators communicate risk by using concepts from the field of public health and safety, expressed in everyday words. Hart and Cooke (2013, p. 99) quote physicians Henderson and Keiding (2005, pp. 705-706) to illustrate how research findings regarding risk can be conveyed: “To illustrate the type of individual risk estimate for survival following diagnosis of non-small-cell lung cancer they thought was warranted by the available scientific data and statistical analyses, they provided the following example: ‘If a group of 90 people like you are followed, research indicates that 30 will die within 4 months, 30 will die between 4 and 11 months, and 30 will live more than 11 months. I do not know which group you will belong to.’” As we move toward comparable ways of communicating sexual-recidivism risk to a fact finder, we will need to introduce some technical concepts. We believe that evaluators

need to understand these concepts, even though we can typically convey the results of a risk assessment using everyday words.

In forensic cases, sexual-recidivism risk assessments are best conceptualized within the context of public health and safety.⁷ Within the field of public health and safety, there are established ways of communicating risk, including the following:

- **Number Needed to Treat (NNT)** is the average number of patients needed to be treated to prevent one bad outcome; that is, the number of patients that need to be treated for one to benefit, compared with a control group in a clinical trial (Cook & Sackett, 1995; Gigerenzer, 2002; Singh, 2013).
- **Positive Predictive Value (PPV)** is the proportion of people predicted to sexually reoffend, who are detected to have sexually reoffended.
- **Negative Predictive Value (NPV)** is the proportion of people predicted not to sexually reoffend, who are not detected to sexually reoffend.
- **Number Needed to Detain (NND)** calculates the average number of individuals judged by a risk-assessment tool to be at high risk of committing a violent act who would need to be detained in order to prevent a single incident of violence from occurring in the community (Buchanan & Leese, 2001; Fleming, 1997; Singh, 2013). Mathematically, the NND is the inverse of the PPV (Buchanan & Leese, 2001; Fleming, 1997).
- **Number Safely Discharged (NSD)** calculates the average number of individuals judged to be at low risk who could be discharged prior to a single violent incident occurring in the community (Fazel, et al., 2012; Singh, 2013).

⁷ For example, Florida's laws regarding the Involuntary Civil Commitment of Sexually Violent Predators are contained within the Public Health section of the Florida Statutes.

NNT. Forensic evaluators are often asked, “Does Mr. X need sex-offender treatment?” That question is comparable to other treatment-related questions, such as, “Does Mr. Y need medicine?” or “Does Mr. Z need chemotherapy?” Such questions entail not only issues about the person, but also about the treatment. Forensic evaluators should not only be prepared to testify about risk (dangerousness), but also about the effectiveness of sex-offender treatment.

Number needed to treat (NNT) is a useful measure of treatment effectiveness (Cook & Sackett, 1995; Gigerenzer, 2002; Singh, 2013). It is the average number of patients needed to be treated to prevent one bad outcome; that is, the number of patients that need to be treated for one to benefit, compared with a control group in a clinical trial. For sex-offender treatment, NNT can be operationalized as the average number of sex offenders needed to be treated to prevent one re-arrest or re-conviction for a sexual offense. Typically, NNT would be calculated from the results of one or more clinical trials. To calculate NNT, we first looked for reviews of relevant clinical trials.

Dennis, et al. (2012, p. 2), recently reviewed sex-offender treatment studies with randomized trials. They found no evidence that psychological treatment reduces sexual recidivism: “Without such evidence, the area will fail to progress. Not only could this result in the continued use of ineffective (and potentially harmful) interventions, but it also means that society is lured into a false sense of security in the belief that once the individual has been treated, [his or her] risk of reoffending is reduced. Current available evidence does not support this belief” (Dennis et al., 2012, p. 2).

One other recent review focused on sexual abusers of children. Långström, et al. (2013) initially reviewed 1,447 abstracts, but found only three randomized control trials, only one of

which involved treatment of adults (Marques, Miederanders, Day, Nelson, & van Ommeren, 2005). That study did not show that sex-offender treatment reduced sexual offending.⁸

Because clinical trials have not shown clear evidence of any treatment effects, we turned to meta-analyses as we continued our quest for data to calculate NNT for sex-offender treatment. Meta-analyses of the effectiveness of sex-offender treatment have generally led to two types of conclusions, both of which are expressed clearly in one recent meta-analysis: “Consistent with previous meta-analyses, the sexual and general recidivism rates for the treated sexual offenders were lower than the rates observed for the comparison groups (based on unweighted averages, 10.9% vs. 19.2% for sexual recidivism). . . . [However,] Reviewers restricting themselves to the better-quality, published studies could reasonably conclude that there is no evidence that treatment reduces sexual offense recidivism. . . . Readers sympathetic to sexual offender rehabilitation⁹ may be content with the encouraging findings from weak research designs; however, skeptics will only be compelled to change their opinions by the strongest possible evidence” (Hanson, Bourgon, Helmus, & Hodgson, 2009, pp. 881, 887). If we use the “encouraging findings from weak research designs,” we can calculate NNT for sex-offender treatment from several recent meta-analyses (all mentioned in Hanson et al., 2009), using a standard NNT calculator.¹⁰ Table 3 presents the results.

Insert Table 3 about here.

⁸ In the one randomized controlled trial of treatment for children with sexual behavior problems, Carpentier, Silovsky, and Chaffin (2006) found that children (aged 5-12) who received 12 sessions of cognitive-behavioral therapy had fewer subsequent detected sexual offenses than children who received play therapy. In the one randomized controlled trial involving treatment of adolescents who had committed sex crimes, Borduin, Schaeffer, and Heiblum (2009) found that adolescents who received traditional treatment provided by experienced, master’s-level therapists who “had been certified as juvenile sexual offender counselors” (p. 29) committed significantly *more* new sex crimes than adolescents treated by graduate students with multisystemic therapy.

⁹ We wonder which readers would be “sympathetic to sexual offender rehabilitation” other than (a) providers of sexual offender rehabilitation and (b) shareholders of certain corporations with government contracts. We would not expect the public to be sympathetic to any particular form of treatment or rehabilitation, whether it involves talking, medicine, etc. The important question is, “Does it work (reduce recidivism)?”

¹⁰ <http://graphpad.com/quickcalcs/NNT1/> Because these data are from meta-analyses rather than an actual clinical trial, we cannot provide realistic confidence intervals.

We could summarize the information in this section as follows: *If we restrict ourselves to the better-quality, published research, we could reasonably conclude that there is no evidence that treatment reduces sexual recidivism. If we base the analysis on encouraging findings from weak research designs, we would expect to have to treat about 13 to 23 sex offenders to prevent one detected sex crime.*

To express that in everyday words, we suggest the following: *We do not know whether sex-offender treatment actually works; that is, whether it actually reduces sexual reoffending. If sex-offender treatment does work, we expect that about 1 in every 13 patients will benefit from treatment; or, perhaps, 1 in every 23 patients will benefit from treatment. In other words, we would expect to have to treat about 13 to 23 sex offenders to prevent one detected sex crime.*

With this research in mind, we return to the aforementioned question that forensic evaluators are commonly asked: “Does Mr. X need sex-offender treatment?” Our conclusion, which we consider to be inescapable, is that any answer that did not provide the above context would likely be misleading, and would thereby violate Specialty Guideline 11.01.

NND and NSD. Some forensic evaluations, including those relevant to civil commitment of sexually violent predators, are conducted in cases in which the fact finder will decide whether or not a person will be confined in order to protect the public. In those cases, the number needed to detain (NND) and the number safely discharged (NSD) are directly relevant to the fact finder’s decision. These calibration performance indicators have recently been clearly presented and described (Singh, 2013, p. 5):

The NND calculates the number of individuals judged by a risk assessment tool to be at high risk of committing a violent act who would need to be detained in order to prevent a single incident of violence from occurring in the community

(Fleminger, 1997). The NSD calculates the number of individuals judged to be at low risk who could be discharged prior to a single violent incident occurring in the community (Fazel, et al., 2012).

Forensic evaluators cannot tell a fact finder whether a particular person will sexually reoffend if not confined. We can tell a fact finder the following, for a person with a Static-99R score of 7: *Mr. X was evaluated using a risk-assessment tool that considers his age, sex-offense history, and other factors. That does not tell us whether Mr. X will sexually reoffend if he is not confined. It does tell us that, in order to prevent one detected sexually violent act within 5 years, we would have to lock up 5 persons with similar risk factors. Also for a 5-year period, we expect that we could release 18 such persons prior to detecting one sexually violent act (see DeClue & Campbell, 2013).*

DeClue and Campbell (2013) calculated calibration performance indicators relevant to detected sexual recidivism during a 5-year fixed follow-up for the samples that the Static-99R developers consider to be roughly representative of all adjudicated sex offenders. That article provides PPV, NPV, NND, and NSD, along with associated confidence intervals, for each Static-99R score.

Singh (2013, p. 13) recognizes that consumers of research or clinical casework who see NND and NSD are directly faced with moral choices: “For example, some may consider the unnecessary detention of, say, five people to prevent the violent behavior of a sixth an appropriate measure to ensure public safety, whereas others may feel that the civil rights of those five unnecessarily detained individuals are of greater importance.” We consider this to be an important advantage for the use of NND and NSD in forensic cases such as those involving civil commitment of SVPs. A proper role for a forensic evaluator is to investigate and then present

data in a way that is quintessentially relevant to the practical and moral decisions to be made by the judge and/or jury. Evaluators need not advocate for or against a particular decision (“better that five people be unnecessarily detained . . .”), but we consider it to be important that the evaluators’ risk communication helps the trier of fact understand the practical impact of the decision to be made.

5. Do clinical adjustments or overrides enhance the accuracy of sexual-recidivism risk predictions?

Specialty Guidelines 2.05, 11.01, and 11.04 are pertinent. 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, this decreases the accuracy of the risk assessment (Gore, 2007; Hanson, 2007; Storey, Watt, Jackson, & Hart, 2012; Vrana, Sroga, & Guzzo, 2008; Wormith, Hogg, & Guzzo, 2012; see also Campbell & DeClue, 2010; DeClue, 2013; Hanson & Morton-Bourgon, 2009). Although it might seem likely that a smart, well-trained expert could use clinical judgment to enhance the accuracy of an actuarial sexual-recidivism risk assessment, no evidence supports that expectation. So far, all of the evidence is to the contrary. Therefore, we recommend that an evaluator who scores an actuarial risk-assessment instrument, but then chooses to express a risk estimate that differs from the results of the actuarial instrument, incurs an affirmative obligation to tell the fact finder that such a practice usually results in less accurate risk predictions. For further discussion, including consideration of possible exceptions, see DeClue (2013).

Whenever an evaluator uses the Static-99R and then discusses risk based on factors outside the Static-99R, we believe that the evaluator has an affirmative obligation to tell the fact

finder that, so far, research shows that consideration of additional factors tends to decrease the accuracy of risk assessments.

6. Is there any objective procedure that can allow me to reliably and validly predict that someone is more-likely-than-not to sexually recidivate if not confined?

Specialty Guidelines 2.05, 10.02, 11.01, 11.02, and 11.04 are pertinent. At present, the answer to this question is no. One might come to this conclusion solely on the basis that predictions at the *individual* level lack sufficient precision to make such a prediction with a reasonable level of accuracy. In fact, we arrive at the same conclusion with regard to predictions at the group level. Much of the rationale is provided above and is summarized here.

- First, consider data regarding the Static-99R FULLPOP comparison group, the group considered to represent the full population of convicted sex offenders. The data in Tables 1 and 2, above, show no levels of the Static-99R that entail a greater-than-50% proportion of detected sexual recidivism. Similarly, there are no PPV values at or above 50% (see Table 2 in DeClue & Campbell, 2013).
- Second, consider that there is, as yet, no empirical support for choosing a specialized comparison group at www.static99.org rather than the representative group. It follows that there is no reliable and valid way to enhance the accuracy of risk assessment by selecting a higher-risk comparison group. Further, there is no empirical evidence that evaluators who choose a specialized comparison group produce more accurate risk assessments than evaluators who consistently use the comparison group representative of the full population of convicted sex offenders.
- Third, although several tools other than the Static-99R have been studied, research has not shown any of them to consistently and effectively identify a subset of sex offenders whose

detected sexual recidivism rate exceeds 50% (see, e.g., Fazel et al., 2012). This holds for both actuarial and SPJ instruments.

- Fourth, there has been no empirical evidence showing that clinical judgment enhances accuracy of prediction beyond that of an actuarial risk-assessment tool.
- Finally, no combination of actuarial tool(s), SPJ tool(s), and/or clinical judgment has been identified that allows evaluators to reliably and objectively identify a subset of sex offenders who are more-likely-than-not to sexually recidivate within any specified time period or within their lifetimes (Abbott, 2009).

We believe that, whenever an evaluator anticipates that his or her testimony might be used to convey that the person is more-likely-than-not to sexually recidivate if he is not confined, the evaluator has an affirmative obligation to tell the fact finder that there is no objective way to predict that an individual is more-likely-than-not to sexually recidivate within a specified time period or within the person's lifetime.

Courtroom Testimony

1. A Cautionary Tale

James Grigson was a forensic psychiatrist who was involved in 167 capital cases. In 1981, it was reported that he was known as Dr. Death: "The prosecution brings Grigson in for a sentencing hearing and asks him about the guilty man's inclination to commit violent crimes in the future. In each of more than 70 such proceedings since 1967, Grigson has testified that the defendant was a 'sociopath' who was dangerous to society,¹¹ and every time, with a single exception, the jury has unanimously voted for the death penalty" (*Time Magazine*, June 1, 1981).

Grigson was eventually expelled from the American Psychiatric Association and the Texas Society of Psychiatric Physicians. It was reported that he violated ethics codes by

¹¹ See Harvey (1997) for discussion of overprecision, the unwarranted overconfidence that one is correct.

“arriving at a psychiatric diagnosis without first having examined the individuals in question, and for indicating, while testifying in court as an expert witness, that he could predict with 100 percent certainty that the individuals would engage in future violent acts” (Bell, 1995).

At the time of his retirement in 2003, Grigson was still highly respected by some, and he was mostly unapologetic. For example, “A leading prosecutor who relied heavily on Dr. Grigson in capital cases called him ‘a fantastic communicator. He is very logical and reasonable and he talks so that jurors can understand completely what his opinion is.’ . . . ‘I’ve made mistakes and I’ve admitted them,’ [Grigson] said, ‘but there is no doubt in my mind that I am 100 percent sure of what my opinion was. I had a lot of experience and a lot of evidence. Otherwise, I sure wouldn’t have done it’” (*Washington Times*, December 20, 2003).

As noted by Richards (2013), there seems to have been a dramatic increase in polarization among experts involved in the assessment of sexual violent predator evaluations. Myside bias occurs when people filter information in a biased manner based on their prior views (Stanovich, West, & Toplak, 2013). Recent research has shown evidence for adversarial allegiance effects in forensic mental health experts (Murrie, Boccaccini, Guarnera, & Rufino, 2013). It follows that we should be aware of sources of bias in our decision making and take steps to minimize their influence (Borum, Otto, & Golding, 1993).

2. Broken Legs

In their 1989 *Science* article, Dawes, Faust, and Meehl noted that, in spite of an increasingly massive and consistent body of evidence, few practitioners seemed to have changed their practice habits (see also Grove, 2005). The first author’s experience in SVP cases over the past 14 years is similar. In their reports and testimony, SVP evaluators routinely use an actuarial instrument (typically the Static-99R these days) and then use their judgment to consider additional

factors, generally resulting in an opinion of increased level of risk, before offering a “professional opinion” regarding the person’s likelihood to sexually reoffend. In doing so, evaluators typically fail to mention that they are using an approach to risk assessment that has been shown to decrease the accuracy of risk predictions, has no known reliability, and fails to produce a probability of reoffense with an associated confidence interval (making it impossible to know the certainty of the risk prediction). Instead, as noted by Abbott (2011), the combination of clinical judgment with actuarial assessment “opens the door to confirmatory and anchoring biases” (p. 228).

Referencing the APA Ethical Principles of Psychologists and Code of Conduct¹² and quoting Paul Meehl, Grove (2005, p. 1236) concludes:

The principle of beneficence therefore generally requires psychologists to choose and use the prediction method that yields the most accurate predictions. . . . “It is therefore foolish, and I would say even immoral, for a trusted . . . expert . . . to employ a method which has a lower hit-frequency than another available method” (Meehl, 1956, p. 163).

As scientist-practitioners, SVP evaluators should apply the results of scientific studies to the cases we evaluate. If the research showed that adjusted-actuarial risk assessments were more accurate than pure-actuarial risk assessments, it would be an evaluator’s responsibility to learn how to perform the best adjusted-actuarial risk assessment possible. But because extant research shows that clinical adjustments do not increase, and often reduce, accuracy of risk assessments, SVP evaluators should generally refrain from using clinical adjustments or overrides in our risk assessments.

Would it be practical for an SVP evaluator to rely solely on an actuarial instrument, no matter what? No. Meehl (1954, 1957) used a hypothetical “broken-leg” example to illustrate an

¹² <http://www.apa.org/ethics/code/index.aspx?item=3>

exception to an actuarial scheme (see also Grove, 2005). A hypothetical professor goes to the movies on Tuesdays with great regularity. Based on extensive observations, an actuarial table is developed showing that, if it is a Tuesday night, the probability that the professor will go to the movies is .9. This Tuesday morning, though, the professor breaks his leg and is put in a hip cast, making it impossible for him to fit in a theater seat. (It is 1954 in this example, and theaters are not yet accessible for people with disabilities.) Based on this rare set of circumstances, a human judge would – and should – override the actuarial-based prediction and predict that the professor will not go to the movies this Tuesday.

Meehl (1957; see also Grove, 2005) highlighted four points about “broken legs” (special features). The following list paraphrases those four guidelines:

1. A broken leg is an objective fact, determinable with high accuracy.
2. The relationship between the broken leg and the predicted event is recognized by all sane people.
3. The broken leg can be considered in isolation (no interaction effects necessary).
4. The relationship between the broken leg and the predicted event is direct, not mediated by theory.

In their article introducing the Static-99, Hanson and Thornton (2000) mention two examples of “special features” of a case that would likely warrant an override of an actuarial instrument: debilitating disease or stated intentions to reoffend. These are rare but realistic features of SVP cases that would satisfy Meehl’s broken-leg guidelines. DeClue (2013, p. 25) concluded, “In the absence of such rare features, SVP evaluators should not adjust or override actuarial-based predictions.”

3. Cross-Examination Regarding an Actuarial-Based Risk Assessment

Recall that we suggested that forensic evaluators could tell a fact finder, regarding the risk posed in a case in which a person had a Static-99R score of 7: *Mr. X was evaluated using a risk-assessment tool that considers his age, sex-offense history, and other factors. That does not tell us whether Mr. X will sexually reoffend if he is not confined. It does tell us that, in order to prevent one detected sexually violent act within 5 years, we would have to lock up 5 persons with similar risk factors. Also, for a 5-year period, we expect that we could release 18 such persons prior to detecting one sexually violent act.*

Now, imagine that the forensic evaluator interviewed the person with a Static-99R score of 7, and the person told the evaluator, “I know I have a problem, and I don’t know how to control myself. If I’m released, I know I’ll do it again. Maybe not the first time I get a chance, but soon and for the rest of my life.” Consider how additional testimony might proceed:

Q: Doctor, you testified that you scored an actuarial risk-assessment tool in this case?

A: Yes.

Q: You testified that most people with risk characteristics similar to those you identified for Mr. X were not detected to sexually recidivate within five years of their release from confinement?

A: Yes.

Q: You testified that, when evaluators have overridden an actuarial risk assessment and based their risk prediction on their clinical judgment, it usually leads to a decrease in the accuracy of the risk assessment, correct?

A: Yes.

Q: Doctor, in this case, you're aware that Mr. X has said that, if he is released, he won't be able to stop himself from committing new sex crimes; in fact, he told you that himself?

A: Yes.

Q: Doctor, do you have any reason to disregard Mr. X's statement, or to discount it?

A: No.

Q: So, in this case, the actuarial tool tells us that most people with risk characteristics like Mr. X's are not detected to sexually recidivate within 5 years?

A: Right.

Q: But Mr. X says he won't be able to stop himself, and you have no reason to disregard or discount that statement?

A: Yes, that is correct.

4. Cross-Examination Regarding a Risk Prediction that a Person is More-Likely-Than-Not to Sexually Reoffend if Not Confined

As mentioned above, it is our understanding that there is no objective way to predict that an individual is more-likely-than-not to sexually recidivate within a specified time period or within the person's lifetime. However, we have heard other experts express the opinion that it is more likely than not that a particular person will sexually reoffend, without clarifying that this

was a subjective opinion rather than one based on scientific research. We suggest the following cross-examination questions for such situations.

Q: Doctor, is it your opinion that if Mr. X is released from confinement, it is more likely than not that he will sexually recidivate?

A: Yes.

Q: How do you know that?

A: . . .

Q: Can you point to any research studies in which men like Mr. X were released from confinement, and we know that more than half of them sexually recidivated?

A: . . .

Q: Describe the research sample. . . . Was that a sample considered to represent the full range of convicted sex offenders or a specialized sample?

A: . . .

Q: What were the particular risk factors or risk characteristics of the people in the group, more than half of whom were detected to sexually reoffend?

A: . . .

Q: Can you point to evidence that shows that this finding generalizes to the full population of convicted sex offenders? That is, among convicted sex offenders, more than half of them with this particular set of risk characteristics will be detected to sexually reoffend if they are not confined?

A: . . .

Q: Doctor, what I'm trying to establish is, can you point to an objective set of risk characteristics or factors and say, for a sample of 1,000 sex offenders, more than 500 of them are detected to sexually recidivate?

A: . . .

Q: If you use one risk assessment tool or five risk assessment tools; if you use an actuarial tool or an SPJ tool; or if you use three actuarial tools and two SPJ tools plus your clinical judgment; can you point to any combination that has been shown to consistently identify a group of sex offenders who, if they're released from confinement, more than half of them are detected to sexually reoffend?

A: . . .

Q: You used the Static-99R in this case?

A: Yes.

Q: Mr. X's score on the Static-99R is 7?

A: Yes.

Q: You're aware that, in samples considered to represent the full population of sex offenders, most people with a Static-99R score of 7 are not detected to sexually reoffend within five years?

A: Yes.

Q: At this time, for that full-population comparison group, there are no reported sexual recidivism rates for longer times, such as 10 or 15 years, or lifetime?

A: Right

Q: Now, you told us about the sexual recidivism rate for the High Risk comparison group, correct?

A: Yes.

Q: And you told us that you believe that Mr. X should be compared to the High Risk comparison group?

A: Yes.

Q: Isn't it true that when an evaluator chooses the High Risk group instead of the full-population group, it guarantees that the risk prediction will be higher?

A: Yes.

Q: Isn't it also true that there is no evidence that evaluators who choose a specialized, High Risk comparison group produce more accurate risk assessments than evaluators who use the full-population comparison group?

A: Yes.

Q: Wouldn't you agree that this is a textbook example of a biased procedure: one that you know is going to produce a higher risk prediction, but you do not know that it will be a more accurate prediction?

A: . . .

Q: Doctor, you told us that, in addition to the Static-99R, you considered additional risk factors to help you decide how likely Mr. X would be to sexually recidivate, correct?

A: Yes.

Q: For that additional consideration, did you use a specified, pre-determined set of risk factors?

A: . . .

Q: Has the set of risk factors that you considered been shown to lead to more accurate risk predictions than, say, an evaluator who would use pure clinical opinion?

A: . . .

Q: Has that set of risk factors been shown to be more accurate, less accurate, or about as accurate as using the Static-99R; or have there been no such studies?

A: . . .

Q: Now let's talk about incremental validity. Can you point to studies that show that using your set of risk factors *in addition to* the Static-99R, leads to a more accurate risk prediction than using the Static-99R alone?

A: . . .

Q: Well, in general, Doctor, are you aware of research that has investigated whether adding professional judgment to an actuarial, sexual-recidivism risk assessment leads to an increase or a decrease in accuracy, compared to the actuarial prediction itself?

A: . . .

Q: Doctor, would it be fair to say that, in order to reach your opinion that it is more-likely-than-not that Mr. X will sexually recidivate if he is not confined, you did not rely solely on objective evidence and procedures? That is, your opinion is based in part on your subjective judgment?

A: . . .

Q: You agree that research has shown that when evaluators use their professional judgment to adjust or override an actuarial risk prediction, their predictions tend to be less accurate than the actuarial risk prediction alone?

A: Yes.

Q: Would it be fair to say that you're asking us to trust your professional judgment in this case, even though research consistently shows that that's a bad idea?

A: . . .

Summary

In this article, we have focused on risk communication in forensic cases regarding sexual recidivism. Specifically, we considered six basic questions regarding the use of the Static-99R in forensic cases. We summarize our answers to these six questions as follows:

1. Should I report risk for sexual recidivism, or risk for detected sexual recidivism? *Detected sexual recidivism.*
2. Should I report individual risk or group risk? *Group.*
3. Which group risk should I report? *In the absence of local norms, the group representing the full population of convicted sex offenders.*
4. How can risk be clearly and accurately conveyed to fact finders? *Using concepts from the field of public health and safety, expressed in everyday words.*
5. Do clinical adjustments or overrides enhance the accuracy of sexual-recidivism risk predictions? *No.*
6. Is there any objective procedure that can allow me to reliably and validly predict that someone is more-likely-than-not to sexually recidivate if not confined? *No.*

References

- Abbott, B. R. (2009). Applicability of the new Static-99 Experience Tables in sexually violent predator risk assessments. *Sex Offender Treatment, 4*, 1-24.
- 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.
- 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 the Law, 39*, 222-230.
- American Psychological Association. (2013). Specialty Guidelines for Forensic Psychologists. *American Psychologist, 68*, 7-19.
- Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Communicating risk for sexual offenders: Risk ratios for Static-2002R. *Sexual Offender Treatment, 7*, 1-12.
- Bell, L. (1995). Groups expel Texas psychiatrist known for murder cases. *Dallas Morning News*. Retrieved September 15, 2013 from <http://www.ccadp.org/DrDeath.htm>

Borduin, C. M., Schaeffer, C. M., & Heiblum, N. (2009). A randomized clinical trial of multisystemic therapy with juvenile sexual offenders: Effects on youth social ecology and criminal activity. *Journal of Consulting and Clinical Psychology, 77*, 26-37.

Borum, R., Otto, R., & Golding, S. (1993). Improving clinical judgment and decision making in forensic evaluation. *Journal of Psychiatry & Law, 21*, 3-76.

Buchanan, A., & Leese, M. (2001). Detention of people with dangerous severe personality disorders: A systematic review. *The Lancet, 348*, 1955-1959.

Cooke, D. J., & Miche, C. (2014). The generalizability of the Risk Matrix 2000: On model shrinkage and the misinterpretation of the area under the curve. *Journal of Threat Assessment and Management, 1*, 42-55.

Carpentier, M. Y., Silovsky, J. F., & Chaffin, M. (2006). Randomized trial of treatment for children with sexual behavior problems: Ten-year follow-up. *Journal of Consulting and Clinical Psychology, 74*, 482-488. doi: [10.1037/0022-006X.74.3.482](https://doi.org/10.1037/0022-006X.74.3.482)

Cook, R. J., & Sackett, D. L. (1995). The number needed to treat: A clinically useful measure of treatment effect. *British Medical Journal, 310*, 452-454. DOI:10.1136/bmj.310.6977.452.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668-1774.

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. (2013). Calibration performance indicators for the Static-99R: 2013

update. *Open Access Journal of Forensic Psychology, 5*, 81-88.

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.

Dennis, J. A., Khan, O., Ferriter, M., Huband, N., Powney, M. J., & Duggan, C. (2012).

Psychological interventions for adults who have sexually offended or are at risk of offending (Review). *The Cochrane Library, 12*. [Cochrane Database Syst Rev](#). 2012 Dec

12;12:CD007507. doi: 10.1002/14651858.CD007507.pub2. Accessible via

<http://www.ncbi.nlm.nih.gov/pubmed/23235646>

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to

predict violence and antisocial behaviour in 73 samples involving 24,827 people:

Systematic review and meta-analysis. *British Medical Journal, 345*, e4692.

DOI:10.1136/bmj.e4692.

Fleminger, S. (1997). Number needed to detain. *British Journal of Psychiatry*, 171, 287.

DOI:10.1192/bjp.171.3.287a

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.

Gigerenzer, G., & Gray, J. A. M. (2011). *Better doctors, better patients, better decisions: Envisioning health care 2020*. Cambridge MA: MIT Press.

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).

Grove, W. M. (2005). Clinical versus statistical prediction: The contribution of Paul E. Meehl. *Journal of Clinical Psychology*, 6, 1233-1243.

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., Babchishin, K. M., Helmus, L., & Thornton, D. (2013). Quantifying the relative risk of sex offenders for Static-99. *Sexual Abuse: A Journal of Research and Treatment, 25*, 482-515.

Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior, 36*, 865-891.

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., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W., Quinsey, V. L., et al. (2002). First report of the Collaborative Outcome Data Project on the effectiveness of psychological treatment of sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 14*, 169-194.

Hanson, R. K., Lloyd, C. D., Helmus, L., & Thornton, D. (2012). Developing non-arbitrary metrics for risk communication: Percentile ranks for the Static-99/R and Static-2002/R sexual offender risk tools. *International Journal of Forensic Mental Health, 11*, 9-23.

- 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., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*, 119-136.
- Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. Static-99 coding rules revised: 2003. Retrieved September 14, 2013, from www.static99.org
- Hart, S. D., & Cooke, D. J. (2013). Another look at the (im)precision of individual risk estimates made using actuarial risk assessment instruments. *Behavioral Sciences and the Law, 31*, 81-102.
- Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Sciences, 1*, 78-82.
- 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, 33*, 1148-1171.

Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2011). 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.

Henderson, R., & Keiding, N. (2005). Individual survival time prediction using statistical models.

Journal of Medical Ethics, 31, 703-706.

Jackson, R. L., & Richards, H. J. (2008). Evaluations for the civil commitment of sexual

offenders. In R. Jackson (ed.), *Learning forensic assessment* (pp. 183-209). New York:

Routledge.

Krauss, D. A., McCabe, J., & Lieberman, J. (2012). Dangerously misunderstood:

Representative jurors' reactions to expert testimony on future dangerousness in a sexually violent trial. *Psychology, Public Policy, and the Law, 18*, 18-49.

Krauss, D. A., & Scurich, N. (2014). The impact of case factors in jurors' decisions in a sexual

violent predator hearing. *Psychology, Public Policy, and the Law, 20*, 135-145.

Långström, N., Enebrink, P., Laurén, E.-M., Lindblom, J., Werkö, S., & Hanson (2013).

Preventing sexual abusers of children from reoffending: Systematic review of medical and psychological interventions. *British Medical Journal*. Published online August 9,

2013. doi: [10.1136/bmj.f4630](https://doi.org/10.1136/bmj.f4630)

- Marques, J. K., Miederanders, M., Day, D.M., Nelson, C., & van Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's sex offender treatment evaluation project (SOTEP). *Sexual Abuse: A Journal of Research and Treatment, 17*, 79-107.
- Meehl, P.E. (1954). Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. Minneapolis: University of Minnesota.
- Meehl, P. E. (1956). Symposium on clinical and statistical prediction: The tie that binds. *Journal of Counseling Psychology, 3*, 163-173.
- Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology, 4*, 268-273.
- Melton, G., Petrila, J., Poythress, N. G., & Slobogin, C. (2007). *Psychological evaluations for the Courts, Third Edition: A Handbook for Mental Health Professionals and Lawyers*. New York: Guilford.
- Miller, H. A., Amenta, A. E., & Conroy, M. A. (2005). Sexually violent predator evaluations: Empirical evidence, strategies for professionals, and research directions. *Law and Human Behavior, 29*, 29-54.

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.

Phenix, A., Helmus, L., & Hanson, R. K. (July 26, 2012). Static-99R & Static-2002R Evaluators' Workbook. Accessed June 27, 2013, from www.static99.org

Richards, H. (2013). Statistical controversies in the Sexually Violent Predator Evaluation (SVPE) Practice Community: Bayesian departure and polarization. In Helfgott, J. B. (Ed.) (2013). *Criminal psychology (Vol. 3): Implications for forensic assessment, policing, and the courts*. Santa Barbara, CA: Praeger.

Rogers, R., & Shuman, D. W. (2005). *Fundamentals of forensic practice: Mental health and criminal law*. New York: Springer.

Schopp, R. F. (2001). *Competence, condemnation, and commitment: An integrated theory of mental health law*. Washington, D.C.: American Psychological Association.

Scurich, N., & John, R. S. (2011). The effect of framing actuarial probabilities on involuntary civil commitment decisions. *Law and Human Behavior, 35*, 83-91.

Scurich, N., & John, R. S. (2012a). A Bayesian approach to the group versus individual prediction controversy in actuarial risk assessment. *Law and Human Behavior, 36*, 237-246.

- Scurich, N., & John, R. S. (2012b). Prescriptive approaches to communicating the risk of violence in actuarial risk assessment. *Psychology, Public Policy, and the Law, 18*, 50-78.
- Slovic, P., Monahan, J., & McGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing the probability versus frequency formats. *Law and Human Behavior, 24*, 271-296.
- Singh, J. (2013). Predictive validity performance indicators in violence risk assessment: A methodological primer. *Behavioral Sciences and the Law, 31*, 8-22.
- Sreevinasan, S., Weinberger, L. E., Francis, A., & Cusworth-Walker, S. (2010). Alice in actuarial-land: Through the looking glass of changing Static-99 norms. *Journal of the American Academy of Psychiatry and the Law, 38*, 400-406.
- 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
- Schwartz, B. K. (1999). The case against involuntary commitment. In A. Schlank & F. Cohen (eds.), *The sexual predator: Law, policy evaluation, and treatment* (pp. 4:1-4:22). Kinston, NJ: Civic Research Institute.

Time Magazine (June 1, 1981). Law: They call him Dr. Death. Retrieved September 15, 2013

from <http://content.time.com/time/magazine/article/0,9171,954817,00.html>

Tully, R. J., Chou, S., & Browne, K. D. (2013). A systematic review of the effectiveness of sex offender risk assessment tools in predicting sexual recidivism of adult male sex offenders. *Clinical Psychology Review, 33*, 287-316.

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.

Washington Times (December 20, 2003). Texas "Dr. Death" retires after 167 capital case trials.

Retrieved September 15, 2013 from

<http://www.washingtontimes.com/news/2003/dec/20/20031220-113219-5189r/?page=all>

Wollert, R. (2006). Low base rates limit expert certainty when current actuarials are used to identify sexually violent predators. *Psychology, Public Policy, and Law, 12*, 56-85.

Wollert, R., & Cramer, E. (2011). The constant multiplier assumption misestimates long-term sex offender recidivism rates. *Law and Human Behavior, 36*, 390-393.

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.

Table 1: Static-99R Scores and Observed Sexual Recidivism within 5 Years
 FULLPOP - Detailed Recidivism Tables Static-99R (October 2009)

Score	N	Detected Recidivism Rate	Detected to Have Sexually Recidivated	Not Detected to Have Sexually Recidivated
-3	40	.000	0	40
-2	65	.000	0	65
-1	260	.027	7	253
0	294	.027	8	286
1	350	.029	10	340
2	350	.040	14	336
3	343	.058	20	323
4	277	.061	17	260
5	193	.145	28	165
6	110	.127	14	96
7	75	.160	12	63
8	28	.286	8	20
9	13	.385	5	8
10	7	.286	2	5
11	1	.000	0	1
Totals	2,406	.060	145	2,261

Table 2: Static-99R FULLPOP
 Estimated 5-Year Sexual Recidivism Rates
 Logistic Regression Estimates

Score	Observed Sexual Recidivism Rate	Predicted Sexual Recidivism Rate	95% Confidence Interval	
-3	0.0	1.2	0.7	2.0
-2	0.0	1.6	1.0	2.6
-1	2.7	2.1	1.3	3.4
0	2.7	2.8	1.8	4.4
1	2.9	3.8	2.5	5.8
2	4.0	5.0	3.4	7.4
3	5.8	6.6	4.6	9.6
4	6.1	8.7	6.1	12.2
5	14.5	11.4	8.2	15.6
6	12.7	14.7	10.8	19.7
7	16.0	18.8	14.0	24.7
8	28.6	23.7	18.0	30.6
9	38.5	29.5	22.8	37.2
10	28.6	--	--	--
11	0.0	--	--	--

Table 3: Number Needed to Treat (NNT) for Sex-Offender Treatment

	Sexual Recidivism		NNT
	Treated	Untreated	
Hanson et al. (2009)	10.9%	19.2%	13
Lösel and Schmucker (2005)	11.1%	17.5%	16
Hanson et al. (2002)	12.3%	16.8%	23