

## Years of Predicting Dangerously

Gregory DeClue, Independent Practice, Sarasota, FL  
[gregdeclue@me.com](mailto:gregdeclue@me.com)

### Abstract

In 2001, Petrila and Otto wrote: "Perhaps most important, there [are] no data on the validity of adjusted actuarial assessment of risk for sexual reoffending, the technique used by almost all professionals who employ actuarial tests in their assessments" (p. 3-8) relevant to the civil commitment of sexually violent predators (SVPs). A decade later, there are data, and the data thus far show that clinical adjustments or overrides reduce the accuracy of actuarial-based risk prediction. What then must we do?

**Keywords:** sexual recidivism, adjusted-actuarial risk assessment, risk assessment, sexually violent predator, SVP

---

In response to public outrage evoked by highly publicized sexual assaults, 20 states and the federal government have enacted sexually violent predator (SVP) statutes (Arizona, California, Florida, Illinois, Iowa, Kansas, Massachusetts, Minnesota, Missouri, Nebraska, New Hampshire, New Jersey, New York, North Dakota, Pennsylvania, South Carolina, Texas, Virginia, Washington, and Wisconsin). These laws mandate a civil hearing to determine whether a previously convicted sex offender should be civilly committed, and confined indefinitely, for treatment. Texas' SVP statute is unique in that it places committed offenders into supervised community settings (Fitch & Hammen, 2003).

This is a form of preventive detention, designed to protect society from persons deemed too dangerous to live unfettered. Civil commitment generally requires a finding such as the following: The respondent has a mental abnormality or personality disorder that makes him likely to engage in new acts of sexual violence if he is not confined for control, care, and treatment. Civil commitment trials typically include expert testimony from psychologists or psychiatrists regarding a person's diagnosis and risk assessment.<sup>1</sup>

States generally recommend or require that evaluators use one or more actuarial risk-assessment tools to assist in assessing risk for sexual recidivism, and clinical consideration of additional potential risk factors. For example, Florida's Sexually Violent Predator Program and Florida Administrative Code 65E-25<sup>2</sup> requires, "The evaluator's clinical opinion shall be the product of clinical judgment guided by the application of assessment instruments helpful in the prediction of sexual offender recidivism" and specifically

---

<sup>1</sup> For more detailed introductions, see Campbell and DeClue (2010a and 2010b) and DeClue and Campbell (2010).

<sup>2</sup> See <https://www.flrules.org/gateway/ChapterHome.asp?Chapter=65E-25>

mentions the Static-99. However, there is no empirical evidence that consideration of additional factors increases the accuracy of the actuarial-based risk assessment. A brief timeline illustrates this.

### **1997**

The Rapid Risk Assessment for Sex Offense Recidivism (RRASOR; Hanson, 1997) was developed to aid in the prediction of sex-offense recidivism using a small number of easily scored variables. Hanson began with a pool of seven items that were correlated at least 0.11 with sex-offense recidivism and were commonly recorded:

- prior sex offenses
- any prior nonsex offenses
- any male victims
- any stranger victims
- any unrelated victims
- never married
- age less than 25 years

Logistic regression analysis was performed for a combined sample of 2,592 offenders. It was found that four variables contributed substantially to the regression equation (that is, added incremental validity to risk prediction) and they were incorporated into the RRASOR:

- prior sex offenses
- any unrelated victims
- any male victims
- age less than 25

It is noteworthy that, even though three variables (any prior nonsex offenses, any stranger victims, and never married) were correlated with sexual recidivism, they did not add incremental validity (increased accuracy) to risk prediction, once the four variables in the RRASOR were considered.

**1998**

Hanson (1998) wrote the following:

Given the current state of knowledge, I believe that there are three plausible approaches to conducting risk assessments: guided clinical, pure actuarial, and adjusted actuarial. In the *guided clinical approach*, expert evaluators consider a wide range of empirically validated risk factors and then form an overall opinion concerning the offender's recidivism risk. In the guided clinical approach, the method for translating the identified risk factors into recidivism rates is not explicitly determined. In contrast, the *pure actuarial approach* evaluates the offender on a limited set of predictors and then combines these variables using a predetermined, numerical weighting system. The *adjusted actuarial approach* begins with an actuarial prediction, but expert evaluators can then adjust (or not) the actuarial prediction after considering potentially important factors that were not included in the actuarial measure (pp. 52-53).

Hanson (1998) recommended caution when attempting to adjust actuarial estimates of recidivism risk.

When the appropriate research has been conducted, what were considered external factors will either be incorporated into an actuarial instrument or will be deemed safe to disregard. . . . Given the poor track record of clinical prediction, evaluators should, nevertheless, be exceedingly cautious about how much they adjust actuarial estimates. Many of the factors that clinicians intuitively believe to be related to sexual offense recidivism, such as denial and verbal statements of treatment motivation, have not been found to predict sexual offense recidivism over the long-term (pp. 65-66).

In their 1998 study reporting correlational data between risk variables and sex offender recidivism, Hanson and Bussière<sup>3</sup> advised:

The predictive accuracy of most of the variables was also small (.10-.20), and no variable was sufficiently related [to recidivism] to justify its use in isolation. It was also unclear how best to combine the variables because their intercorrelations were unknown and would be expected to be rather high for certain variables (e.g., young and unmarried). Consequently, we

---

<sup>3</sup> This is the meta-analysis mentioned in the discussion of the RRASOR (above). Although the RRASOR was published in 1997 and the meta-analysis was published in a peer-reviewed journal in 1998, data from the meta-analysis were originally made available in 1996. See <http://www.publicsafety.gc.ca/res/cor/rep/1996-04-psorma-eng.aspx>

do not recommend simply summing the items, using either unit weights or weights inferred from the [correlational] tables, to create risk scales. . . . Nevertheless, our results could be used to identify the factors worth considering in risk assessments (p. 358).

As described above, the meta-analysis was used to identify factors worth considering, which helped shape the RRASOR.

## 2000

The Static-99 was developed by combining the RRASOR with another instrument, the SACJ-Min (Hanson & Thornton, 2000). In addition to the four items from the RRASOR (see above), the Static-99 includes the following six additional items:

- never married<sup>4</sup>
- noncontact sex offenses
- stranger victims<sup>5</sup>
- current nonsexual violence
- prior nonsexual violence
- four or more sentencing dates

Hanson and Thornton 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).<sup>6</sup> That is, adding six new items, which more than doubled the total number of items, increased overall accuracy of sex-offense risk by a small amount. The authors also noted, “For the prediction of sex-offense recidivism, Static-99 is clearly more accurate ( $r = 0.33$ ) than unstructured clinical judgment (average  $r = 0.10$ )” (p. 131).

Hanson and Thornton (2000) speculated that adjusting actuarial assessments might eventually be shown to enhance accuracy of risk prediction, but they urged caution about any such adjustments in the absence of evidence that adjustments improve accuracy (p. 132):

Static-99 does not claim to provide a comprehensive assessment, for it neglects whole categories of potentially relevant variables (e.g., dynamic factors). Consequently, prudent evaluators would want to consider whether there are external factors that warrant adjusting the initial score or special features that limit the applicability of the scale (e.g., a debilitating disease or

<sup>4</sup> This item did not add incremental validity to the four items that made it into the RRASOR (see above).

<sup>5</sup> This item did not add incremental validity to the four items that made it into the RRASOR (see above).

<sup>6</sup> Perhaps “tiny and insignificant” would be more fitting than “relatively small.” According to their Table 4 on page 126, Receiver Operating Characteristic (ROC) Area for RRASOR was 0.68, with a 95% confidence interval (CI) of 0.65 to 0.72. ROC Area for Static-99 was 0.71, with a 95% CI of 0.68 to 0.74. (Hanson & Thornton, 2000).

stated intentions to reoffend). Given the poor track record of clinical prediction, however, adjustments to actuarial predictions require strong justifications. In most cases, the optimal adjustment would be expected to be minor or none at all.

## **2001**

Petrila and Otto (2001) wrote: "Perhaps most important, there [are] no data on the validity of adjusted actuarial assessment of risk for sexual reoffending, the technique used by almost all professionals who employ actuarial tests in their assessments" (p. 3-8).

## **2002**

Hanson (2002, p. 100) wrote, "Much more research is required before adjustments to established actuarial measures using static factors can be done with any confidence."

## **2005**

Hanson and Morton-Bourgon (2005) completed a new meta-analysis. They commented on their data regarding risk factors and recidivism (p. 1159):

Readers will note, however, that the predictive accuracy of most of the characteristics was small. Consequently, prudent evaluators need to consider a range of potential risk factors in an overall evaluation. The best methods for combining risk factors into an overall evaluation remain an active topic of scientific debate.

## **2009**

The Static-99 was revised in October 2009. The revised instrument, Static-99R, has the same items, 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, 2011). It was also recognized that sexual recidivism rates are much lower than in years past. The developers of the instruments now recommend the Static-99R rather than the Static-99 for all purposes.

In 2009, Hanson and Morton-Bourgon completed yet another meta-analysis. This time they found three studies that directly compared the accuracy of pure-actuarial risk assessment to adjusted-actuarial risk assessment, for sexual recidivism. They wrote (p. 7):

Three studies examined the difference between actuarial scores and adjusted actuarial risk ratings (Gore, 2007; Hanson, 2007; Vrana, Sroga, &

Guzzo, 2008). In these studies, evaluators were required to complete an actuarial risk tool and then were allowed to adjust the final risk rating on the basis of factors external to the actuarial tool. All three studies were prospective, and evaluators completed the ratings as part of their routine procedures. 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.

## 2010

Campbell and DeClue (2010a, p. 75) looked at the publicly available data regarding those same three studies and reached a more modest conclusion: “Based on available data, at its best, AAA<sup>7</sup> neither increases nor decreases the accuracy of actuarial classification. At its worst, AAA dilutes actuarial accuracy.”

How do adjustments or overrides to actuarial risk assessments dilute accuracy? One example is found in Gore’s (2007) dissertation. She found that clinical overrides that increased predicted risk resulted in 4 more true positives (people rated at high risk, who actually sexually recidivated) but at the cost of 75 fewer true negatives (people rated as low risk, who actually did not sexually recidivate).<sup>8</sup>

## 2011

Montaldi (2011) mentioned, “Given decreased base-rates over the past 20 years, the most accurate method now may be to just use the overall (low) reconviction base-rate and predict non-reconviction for every offender. We would have false negative errors but perhaps fewer errors overall.”<sup>9</sup>

## 2012

Two more studies have addressed the accuracy of pure-actuarial risk assessment versus adjusted-actuarial risk assessment for sexual recidivism. In one study, Storey, Watt, Jackson, and Hart (2012) found that clinical adjustments or overrides of the Static-99 decreased the accuracy of risk prediction: “In 30 cases, clinicians used discretion to ‘override’ or adjust the Static-99 ratings when making final risk judgments, but the

---

<sup>7</sup> Adjusted-actuarial assessment.

<sup>8</sup> “When psychologists’ overrides were attended to in only one direction (e.g. upward overrides) and ignored in the other direction (e.g., downward overrides), the adjusted risk levels were significantly more accurate in the direction in which overrides were permitted (e.g., identification of true positives). However, this improvement in accuracy was achieved at a statistically significant and extreme cost: zero accuracy in the other direction (e.g., identification of true negatives)” (Gore, 2007, pp. 51-52).

<sup>9</sup> Dr. Montaldi is the director of Florida’s Sexually Violent Predator Program.

predictive validity of the clinical adjusted ratings was worse than that of the original Static-99 ratings made by clinicians” (p. 1). “The clinical override scores were less predictive of sexual recidivism than the scores without overrides. . . . Examination of the Hazard Ratios for the two sets of ratings indicated that 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” (p. 8).

Storey et al. reported that their results comport with prior research (as noted above): “In addition to altering risk determinations, clinical overrides also decreased the predictive validity of Static-99 scores. This decrease was to be expected based on the recommendations of most proponents of actuarial assessment instruments not to override scores, as well as the premise on which actuarial instruments are based, namely that optimal empirically based algorithms as opposed to unstructured clinical judgment are the best predictors of future violence. In addition, other studies have also found that clinical overrides made to actuarial scores decrease predictive validity (Gore, 2007; Hanson, 2007; Vrana, Sroga, & Guzzo, 2008)” (p. 9). See also Hanson, Harris, Scott, and Helmus (2007).

Storey et al. concluded, “Clinical judgment reduced the predictive accuracy of the Static-99 in our study. . . . On the basis of our findings, 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 overriding Static-99 scores under any circumstances” (pp. 10-11).

In another study published in 2012, 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 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).

## Summary

Two somewhat contradictory recommendations emerge from this timeline, separated by a dozen years of research and practice:

- Hanson & Thornton (2000): Prudent evaluators would want to consider whether there are external factors that warrant adjusting the initial score or special features that limit the applicability of the scale (e.g., a debilitating disease or stated intentions to reoffend). Given the poor track record of clinical prediction, however, adjustments to actuarial predictions require strong justifications. In most cases, the optimal adjustment would be expected to be minor or none at all.
- Storey et al. (2012): 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 overriding Static-99 scores under any circumstances.

SVP evaluators now have five research studies directly comparing pure-actuarial to adjusted-actuarial risk assessment for sexual recidivism. Overall, the evidence does not show that adjusting or overriding the results of an actuarial instrument increases the accuracy of risk assessment.<sup>10</sup> On the contrary, there is evidence that adjustments or overrides often decrease accuracy. These five research studies lead to the conclusion that evaluators have *not* been prudent. In spite of Hanson and Thornton's guidelines (be prudent, make only minor adjustments or none at all, only make adjustments in the face of special features such as debilitating disease or stated intentions to offend), clinicians have sometimes chosen not to rely on the results of the actuarial instrument, even in the absence of special circumstances. The result has been a decrease in accuracy. Available research does not support the use of professional judgment to adjust or override actuarial-based risk assessment of sexual recidivism.

### What Then Must We Do?

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). My 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 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).

---

<sup>10</sup> For a possible, limited exception, see Wormith et al. (2012, p. 1530): "Our findings suggest that the use of the override function to lower risk level, as rare as that was, may indeed be an appropriate procedure." This has not been a consistent finding across studies.



Referencing the APA Ethical Principles of Psychologists and Code of Conduct<sup>11</sup> 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 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 re-offend. These are rare

---

<sup>11</sup> <http://www.apa.org/ethics/code/index.aspx?item=3>

but realistic features of SVP cases that would satisfy Meehl's broken-leg guidelines. In the absence of such rare features, SVP evaluators should not adjust or override actuarial-based predictions.

Received November 8, 2012; revision submitted January 8, 2013; accepted January 12, 2013.

## 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.
- Campbell, T. W., & DeClue, G. (2010a). Flying blind with naked factors: Problems and pitfalls in adjusted-actuarial sex-offender risk assessment. *Open Access Journal of Forensic Psychology*, 2, 75-101.
- Campbell, T. W., & DeClue, G. (2010b). Maximizing predictive accuracy in sexually violent predator evaluations. *Open Access Journal of Forensic Psychology*, 2, 148-232.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668-1774.
- DeClue, G., & Campbell, T. W. (2010). Still maximizing accuracy in sexually violent predator evaluations. *Open Access Journal of Forensic Psychology*, 2, 322-336.
- Fitch, W. L., & Hammen, D. A. (2003). The new generation of sex offender commitment laws: Which states have them and how do they work? In B. J. Winick & J. Q. LaFond (Eds.) *Protecting Society from sexually dangerous offenders*. Washington, DC: American Psychological Association.
- 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. (1998). What do we know about sex offender risk assessment? *Psychology, Public Policy, and Law*, 4, 50-72.
- 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. (2002). Introduction to the Special Section on dynamic risk assessment with sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, 14, 99-101.
- 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., Harris, A. J. R., Scott, T.-L., & Helmus, L. (2007). Assessing the Risk of Hanson, R. K., & Morton-Bourgon, K. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, 73, 1154-1163.
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, 21, 1-21.
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24, 119-136.
- 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*.
- 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.
- Montaldi, D. F. (2011). Probability and sexual offenders: Some preliminary remarks. *The Forum*, XXIII (2), 20-32.
- Petrila, J., & Otto, R.K. (2001). Admissibility of expert testimony in sexually violent predator proceedings. In A. Schlank (Ed.), *The sexual predator: Legal issues, clinical issues, special populations - Volume II*. Kingston, NJ: Civic Research Institute.

- Storey, J. E., Watt, K. A., Jackson, K. J., & Hart, S. D. (published online February 17, 2012). Utilization and implications of the Static-99 in practice. *Sexual Abuse: A Journal of Research and Treatment*.  
<http://sax.sagepub.com/content/early/2012/02/16/1079063211423943>
- 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.
- 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.