Still Maximizing Accuracy in
Sexually Violent Predator Evaluations[1]

Gregory DeClue, Ph.D., ABPP (forensic), Sarasota, Florida
gregdeclue@me.com

Terence W. Campbell, Ph.D., ABPP (forensic), Sterling Heights, Michigan

Abstract

We respond to eleven points raised by Wilson and Looman (2010) regarding our article, "Maximizing Predictive Accuracy in Sexually Violent Predator Evaluations" (Campbell & DeClue, 2010b). We summarize our recommendations regarding risk assessment and risk communication. We continue to recommend that evaluators who use actuarial instruments should inform the trier of fact how to maximize the accuracy of predictions based on those instruments.

Keywords: sexual recidivism, risk assessment, actuarial risk assessment, predictive accuracy, base rates, cutoff scores, Static-99, Static-99R, Static-2002, Static-2002R, sexually violent predator (SVP), forensic psychology

---

We appreciate the opportunity to respond to Wilson and Looman's (2010) critique of our article about how to maximize predictive accuracy in sexually violent predator evaluations. Here, we respond to eleven points raised by Wilson and Looman. We will respond to the items point by point in the order in which they appear in Wilson and Looman's article, and then present a general discussion.

Point by Point

In addition to commenting about our article, Wilson and Looman provide a historical and practical summary about actuarial risk prediction and related matters. We agree with much of what they present. Our focus here is on points of disagreement.

1. In the abstract, Wilson and Looman write (p. 306), "In their paper 'Maximizing Predictive Accuracy in Sexually Violent Predator Evaluations,' Campbell and DeClue argue that the predictive value of the Static-99 and its related measures is suspect."

We prefer to frame our paper as a consideration of how to maximize predictive accuracy in SVP evaluations, rather than as an argument about the value of particular measures.

---

[1] Associate Editor's Note: The first author of this manuscript is the journal's editor and publisher. I supervised the editorial process for this article.

Indeed, early in our article (page 149) we present the following five propositions (see the original for citations):

> In published, peer-reviewed research, overall, on average,
>
> 1. The instruments in the Static-99 group are about as accurate as, or more accurate than, other existing actuarial instruments.
>
> 2. There is no evidence that adding an additional actuarial instrument improves accuracy, compared to one of these instruments alone.
>
> 3. There is no evidence that structured professional judgment (SPJ) is more accurate than the use of an instrument from the Static-99 group.
>
> 4. There is no evidence that clinical adjustment to one of these instruments improves accuracy.
>
> 5. There is no evidence that we can currently predict an individual's risk to sexually re-offend better than a prediction derived from knowing the sexual-recidivism base rate and using one of the instruments in the Static-99 group.
>
> It is important for anyone using or interpreting one of the instruments in the Static-99 group to understand how accurate the instruments are. The accuracy of these instruments is especially important in light of the above propositions. How accurate are current sex-offender risk assessments? They are no more accurate than the instruments in the Static-99 group.

We see this as an important, overriding contrast. Wilson and Looman address *whether* sexual recidivism can be accurately predicted: "At the heart of the matter is whether or not risk to reoffend can be accurately predicted." In contrast, our primary focus in our paper was *how* to maximize the accuracy of risk assessments and risk communications.

2. In describing the development of risk-assessment tools, Wilson and Looman write (p. 308), "Standardization of the measure requires scoring a sample of individuals on the scale, following them for a period while they are exposed to risk, and then statistically analyzing the outcome data."

Yes, but some tools are developed retrospectively, using data that have already been collected. We mention this solely to clarify a minor point.

3. In discussing Florida's SVP process, Wilson and Looman note that only about 3.5 percent of sex offenders are recommended for civil commitment trials. They write (p. 309), "As it makes some sense that these civil proceedings would be

used with only those offenders who pose the greatest risk to the public, the logical supposition is that those who pose the greatest risk are exceedingly rare."

Even after reading this sentence several times in context, we are unsure whether it is at heart a tautology or an unsupported assumption. Perhaps they are expressing that 3.5% is a small number. If so, fine. However, if they are suggesting that those 3.5% are in fact the most dangerous offenders, it is worth mentioning that we do not know that. Although Florida's Sexually Violent Predator Program (for example) aspires to identify the most dangerous sexual offenders and recommend them for civil commitment, we do not have empirical data regarding the level of success at that endeavor.

4. Wilson and Looman write (p. 311), "Equating risk assessment with medical diagnosis, as Campbell and DeClue attempt to do, is patently inappropriate." They quote Hanson and Howard (2010, p. 276), who wrote: "In contrast to diagnoses, risk assessments estimate the likelihood of an event that has not yet happened, and may never happen. There (sic) are inherently stochastic,[2] and the future outcome can only be estimated with a certain probability."

This is central to our article, and we will address it at some length. We are not alone in construing risk prediction as subject to binary classification analysis. In our article, we mention three strong recommendations that risk assessments are subject to binary classification analysis (p. 153):

> Addressing the accuracy of violence predictions, Hart, Webster and Menzies (1993, p. 698) recommended: "To help prevent future misinterpretations and to facilitate inspection and (re-) analysis by readers, we recommend that journal editors require authors to report in their manuscript the raw data for any 2 X 2 analyses. Such data are easily presented in the form of text or tables."

> Consistent with the positions of Hart et al. (1993), Serin and Brown (2000, p. 263) list as Commandment 8 in their "10 Commandments of Risk Assessment": "Thou shalt know thy false positive and false negative rates for specific cut offs." Additionally, Craig, Browne, and Stringer (2004, p. 8) advise, "Any risk classification compares prediction with actual outcome using 2 X 2 contingency tables."

---

[2] We briefly address two minor points here. First, we do not equate risk assessment with medical diagnosis. Second, even if we did, some diagnostic medical procedures, including EKG, EEG, blood pressure, and temperature, are also considered to be stochastic processes. For a basic description, see http://en.wikipedia.org/wiki/Stochastic_process The fact that binary classification is useful in some medical testing does not support a claim that it is an inappropriate way to assess the accuracy of risk predictions.

Another prominent writer and researcher regarding risk assessment wrote the following in 1998:

> Before reviewing the research evidence on risk assessment, some basic comments on the measurement of predictive accuracy are in order. The evaluator's task is to correctly identify those offenders who will reoffend and those who will not. Because prediction is never expected to be exact, two types of errors are always possible. The evaluator may incorrectly believe that an offender will recidivate or incorrectly believe that the offender will not recidivate. Each case fits into only one of the four cells of the 2 X 2 prediction table: (a) dangerous/reoffends, (b) safe/reoffends, (c) dangerous/does not reoffend, and (d) safe/does not reoffend (see Figure 1). Although the accuracy of 2 X 2 tables is often indexed by the "percentage correct" ([a + d] / [a + b + c + d]), the percentage correct is highly influenced by the recidivism base rate ([a + b] / [a + b + c 4- d]). If few offenders recidivate, as is often the case in sexual offender outcome studies, then the simplest way to maximize the percentage correctly classified is to predict that no one will reoffend. Consequently, researchers prefer to use statistics that measure the degree of association between the prediction and the outcome.

That quote is from page 53 of Hanson's (1998) article in *Psychology, Public Policy, and Law,* "What do we know about sex offender risk assessment?" It is noteworthy that the following conclusions, which could be drawn from Campbell and DeClue (2010b), were explicitly stated by Hanson in 1998 as "some basic comments on the measurement of predictive accuracy":

- "The evaluator's task is to correctly identify those offenders who will reoffend and those who will not."

- "Each case fits into only one of the four cells of the 2 X 2 prediction table."

- "If few offenders recidivate, as is often the case in sexual offender outcome studies, then the simplest way to maximize the percentage correctly classified is to predict that no one will reoffend."

Thus, in construing risk assessment as subject to binary classification analysis, we are in concurrence with Hanson (1998), as well as Hart et al. (1993), Serin and Brown (2000), and Craig et al. (2004). For context, consider that all of the following are binary decisions:

- An SVP evaluator's decision in preparing a written report, whether the person is "likely,"[3] "more likely than not,"[4] etc., to engage in sexual violence.

---

[3] For example, in Florida. See Florida Statutes, Chapter 394.

- The decision of a state's Sexually Violent Predator Program whether to recommend that the Office of the State Attorney file a petition for civil commitment.

- The decision of the Office of the State Attorney whether to file a petition for civil commitment.

- A judge's decision whether to find probable cause to confine the respondent and set the case for trial.

- A judge's or jury's decision whether to find that the person meets criteria for civil commitment.

5. Wilson and Looman write, "Campbell and DeClue . . . go on to make the case that, with the new recidivism norms discussed above, the greatest predictive accuracy is achieved when predicting that no one will re-offend . . ." (p. 311).

That accurately reflects the empirical results of most of our analyses. We remind the reader that this is quite similar to what Hanson wrote in 1998 (p. 53): "If few offenders recidivate, as is often the case in sexual offender outcome studies, then the simplest way to maximize the percentage correctly classified is to predict that no one will reoffend." This is not new. If, in 2010, it is controversial, perhaps that is in part because, as Hanson (1998, p. 53) wrote in the next sentence, "Consequently, researchers prefer to use statistics that measure the degree of association between the prediction and the outcome."

As researchers compare the accuracy of various risk-assessment instruments, and as evaluators choose which instrument to use, it is useful to consider various "statistics that measure the degree of association between the prediction and the outcome." When it comes to interpreting the meaning of a score on a particular risk-assessment instrument, evaluators—and judges, juries, and others—get more directly useful information from the positive predictive value (PPV) and its confidence interval. That brings us to another point mentioned by Wilson and Looman.

6. Wilson and Looman write (p. 311), "Campbell and DeClue . . . go on to make the case that . . . the confidence interval which can be ascribed to individual risk predictions (based on group data) is so broad as to be essentially meaningless."

In our article, we recommend the use of confidence intervals in reporting risk, but we do not address ascribing confidence intervals to individual risk predictions. We invite the reader to compare brief sections from two papers published in 2010. The first section is from Hanson and Howard (2010, p. 279, citation omitted, emphasis added):

---

[4] For example, in Missouri. See Missouri Statutes, Chapter 632.

How then, should the accuracy of risk assessments be reported? We believe that one plausible approach to reporting predictive accuracy is to *report the confidence interval* for the estimate derived from the group that the individual most closely resembles. Evaluators will also need to make a separate, non-quantitative judgment.[5]

The second brief section is from our article (Campbell & DeClue, 2010b, p. 166):

We recommend a straightforward procedure for evaluators to follow in conducting the actuarial portion of sexual-recidivism risk assessments:

1. Identify the most appropriate risk-assessment instrument for the population.

2. Identify (or estimate) the local base rate.

3. Compute and report the PPV, with confidence intervals, for that score on that test with that base rate.

Because test utilities, including PPV, of the Static-99 vary along with variations in base rates, evaluators need to know (or reasonably estimate) their local base rate in order to interpret a Static-99 score.

Throughout our article, when we discuss confidence intervals, it is about confidence intervals around PPV. We consistently address group data, not individual confidence intervals. The example in our Appendix A, involving the Static-2002R, is for group data: For the 526 subjects in the Static-2002R Routine Sample, the 5-year detected sexual recidivism base rate was .053. Using a cutoff score of 11 or higher, the PPV is 0.5, with a 95% confidence interval of about .02 to .97. See our Tables 13 (from page 170) and B-23 (from page 212).

---

[5] Our decision to not quote Hanson's discussion of the non-quantitative judgment is not an indication of whether we think it is important. Here, our focus is on what Hanson calls quantitative judgment and what we call the actuarial portion of a risk assessment.

Table 13[6]
Static-2002R Scores and Observed Sexual Recidivism within 5 Years
Routine Sample

Score	N	Detected Recidivism Rate	Detected to Have Sexually Recidivated	Not Detected to Have Sexually Recidivated
-2	8	.000	0	8
-1	16	.000	0	16
0	36	.000	0	36
1	48	.021	1	47
2	57	.018	1	56
3	70	.029	2	68
4	79	.038	3	76
5	69	.072	5	64
6	67	.060	4	63
7	39	.128	5	34
8	21	.143	3	18
9	12	.250	3	9
10	2	.000	0	2
11	2	.500	1	1
12	0	.000	0	0
13	0	.000	0	0
Totals	526	.053	28	498

Table B-23
Accuracy Levels for Static-2002R, Cutoff of 11 or Higher, 5-Year Follow-up
Routine Sample

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	1	1	2
Not Predicted to Reoffend	27	497	524
Totals	28	498	526

Sample BR = .053, TPR (Sensitivity) = .04, FPR = .002, Specificity = .998
Overall Accuracy at this BR = .9468, PPV at this BR = .50, NPV at this BR = .95

Although the 95% confidence interval for PPV is very wide, it is not an artificial statistic derived from estimating the risk for an individual person. It addresses the degree of

---

[6] For the readers' convenience, these tables are numbered the same as they were in our previous article.

statistical confidence associated with the conditional probability that persons in that sample with a score of 11 or higher were detected to sexually reoffend within five years.

For perspective, consider the negative predictive value (NPV) for the same sample with the same cutoff score. NPV is .95, with a 95% confidence interval of about .92 to .97. Again, see our tables B-23 and 13. Also, see the discussion in Appendix A of Campbell and DeClue (2010b).

Why is it that the confidence interval around PPV (predictions that persons would sexually reoffend) is so large (.02 to .97) and the confidence interval around NPV (predictions that persons would not sexually reoffend) is so small (.92 to .97)? We believe that the key lies in the data, not in our methods of computing confidence intervals. As can be seen in Table 13, at a cutoff score of 11, half of the people sexually reoffend, and half do not. At all lower cutoff scores, most of the people are not detected to sexually reoffend.

As can be seen in Table B-23, which uses a cutoff score of 11, we "predict" that everyone with a Static-2002R score of 11 or higher will be detected to sexually reoffend. There are two such subjects. One of them was detected to sexually reoffend, and the other one was not. The reason the PPV is .5 is that half (1 of 2) people in that sample were successfully "predicted" to sexually reoffend.

It can also be seen in Table B-23 that, still using a cutoff score of 11, we "predict" that everyone with a score below 11 will not be detected to sexually reoffend. There are 524 such subjects, and 497 (95%) of them were not detected to sexually reoffend. NPV is .95.

For this group data, a key reason that the confidence interval for PPV is so wide is that the "prediction" is based on 2 subjects. In contrast, for NPV, the "prediction" is based on 524 subjects, and the confidence interval is much tighter.

To illustrate the same point about uncertainty using language other than "confidence interval" (and relying less on math), consider an altered version of a recommendation from Hanson and Howard (2010, p. 280) for how an evaluator could express findings from the quantitative stage of a risk assessment. The altered version is "Previous studies have found that 50% of offenders with similar characteristics as Mr. X were reconvicted of a sexual offense within 5 years."[7] We believe that it would be misleading for an evaluator to say or write, "In a study of 526 subjects, people who scored on the Static-2002R as Mr. X did, had a 50% chance of reoffending within 5 years," at least unless the evaluator clarifies that claim. "In that study, there were two people who

---

[7] The original was "Previous studies have found that 20% of offenders with similar psychological characteristics as Mr. X were reconvicted of a sexual offense within 5 years." We changed "20%" to "50%" to maintain consistency with the example we have been considering. We deleted "psychological" because some of the variables (e.g., age, number of prior convictions) are usually not called "psychological" characteristics.

scored on the Static-2002R like Mr. X did. One of them was detected to sexually reof-fend within the five-year period, and the other one did not."[8]

7. Wilson and Looman write (p. 312, emphasis in original), "Simply put, the Static-99 and its progeny were not designed to facilitate answering of the 'more likely to than not' question faced by SVP evaluators, *no matter how much they may want them to. . . .* Campbell and DeClue appear to argue that we must be able to sepa-rate offenders into finite groups–those who will and those who will not reoffend."

Florida's Administrative Code can help us consider these statements. Use of the Static-99 is mandated by Florida's Sexually Violent Predator Program and Florida Administra-tive Code 65E-25.[9] Prior to October 2009, the program employed the Static-99. Since that time, the program has changed to the Static-99R. Florida Administrative Rule 65E-25.001(2)(b) states, "*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. Each clinical evaluation shall include the Static-99, if suf-ficient information is available to score that instrument. Evaluators may, in addition to the Static-99, use other assessment instruments as appropriate. *The clinical evaluation shall result in a written report that addresses, at a minimum, whether the evaluated indi-vidual suffers from a mental abnormality or personality disorder that makes the person likely to engage in acts of sexual violence if not confined in a secure facility for long-term control, care, and treatment."* [emphasis added]

The two italicized sections address the points raised just above by Wilson and Looman. Truncating the last sentence quoted from Florida's Administrative Code, the Code requires that the clinical evaluation *shall* result in a written report that addresses . . . whether the evaluated individual is likely to engage in acts of sexual violence if not con-fined. So, evaluators must be able to—or at least must attempt to—separate offenders into finite groups: those who are likely to engage in acts of sexual violence, and those who are not likely to do so. Disregarding for the moment the distinction between "likely" and "more likely than not," the Administrative Code demands that evaluators separate referred persons into two finite groups. The earlier italicized sentence requires that the evaluator's clinical opinion *shall* be guided by the assessment instruments, which *shall* include the Static-99. To be sure, Florida's Administrative Code does not assert that evaluators shall be correct every time, but the Code does require that the evaluator

---

[8] An evaluator could present a risk prediction based on more subjects if a lower cutoff score were used. For example, with a cutoff score of 7, the evaluator could report the detected sexual recidivism rate for 64 subjects. As can be seen in Table 13 above and in Table B-21 on page 211 of our prior article, 12 of those 64 people (19%) were detected to sexually recidivate within the 5-year follow-up period. The greater number of subjects would lead to a somewhat narrower confidence interval for PPV (that is, more confidence in the accuracy of the assessment).

[9] See https://www.flrules.org/gateway/ChapterHome.asp?Chapter=65E-25

shall be guided by the Static-99, and shall address whether the person is likely to engage in acts of sexual violence if not confined.[10]

In considering Wilson and Looman's statement that "The Static-99 and its progeny were not designed to facilitate answering of the 'more likely to than not' question faced by SVP evaluators, we look to the "Static-99 Coding Rules Revised, 2003" at www.static99.org. The following is clearly stated on page 3: "This instrument provides explicit probability estimates of sexual reconviction." For the samples in research groups, scores on Static-99 (or related instruments) can be used to calculate an explicit probability of sexual reconviction. That is the positive predictive value (PPV), with confidence intervals.

8. Wilson and Looman write, "All in all, attendance to 'in-treatment change'– throughout intensive institutional treatment to community-based (and, perhaps, lifelong) maintenance–is probably a more informative endeavor than simply measuring who did or did not commit a new sexual offense. This is particularly true for SVPs . . ." (p. 312, citations omitted).

We can imagine that there may be some situations in which attendance to treatment change is more important than measuring who did or did not commit a new sexual offence. However, neither of us has ever been involved in such a situation. We have no further comment about that.

9. Wilson and Looman write (pp. 313-314, citations and caveats omitted), "There is a certain degree of face validity associated with the process. For example, most evaluators will agree that a person who scores a "3" on the Static-99 and who is also phallometrically judged to be pedophilic warrants greater attention than a person with a similar score who does not show pedophilic sexual arousal. Similarly, a person with Static-99 score of "6" combined with a Psychopathy Checklist-Revised (PCL-R; Hare, 2003) score of "30" causes more concern than someone with a similar Static-99 score and a PCL-R score of "18."

We addressed similar examples in a prior paper, "Flying blind with naked factors: Problems and pitfalls in adjusted actuarial sex-offender risk assessment" (Campbell & DeClue, 2010a), so we will be brief here. It is telling that Wilson and Looman present the impact of such examples in terms of "most evaluators agree," "warrants greater attention," and "causes more concern," rather than mentioning empirical findings. In addition, such comparisons are artificial, in that no judge or jury is presented with two sex offenders and instructed to civilly commit one of them and let the other one go. Instead, in a typical SVP case, a judge or jury considers the fate of one person at a time.

---

[10] In practice, Florida's Sexually Violent Predator Program requires contract evaluators to include a yes (meets criteria for civil commitment) or no "bottom line" opinion in every report.

To take the first example first, an evaluator might be tempted to tell the trier of fact that the evaluee should be considered more dangerous than most people with a Static-99 score of 3, because this person is known to have shown a physiological attraction to children. As we mentioned in the "Flying blind" paper, there are at least two major problems with such an interpretation. First, the Static-99 includes items that are intended to tap paraphilia and antisociality, and are easily measurable from the type of data found in prison files. Therefore, it can be a form of double dipping to treat paraphilia and antisociality/psychopathy as if they were risk factors that are independent of the items on the Static-99.

Second, in the absence of data showing that a variable adds incremental validity to a ten-item risk-assessment instrument, is it clinical judgment or a guess whether and to what extent the additional variable increases risk beyond what would be expected based on the actuarial instrument alone? We remind readers that when six items were added to the four-item Rapid Risk Assessment for Sex Offense Recidivism (RRASOR; Hanson, 1997) to form the Static-99, there was only a small increase in accuracy (AUC increased from .68 to .71; see Table 4, page 126 of Hanson & Thornton, 2000). Without supporting data, it would be precarious to predict that adding one item to the resulting ten-item scale (Static-99) would increase the accuracy of risk predictions.

The other example Wilson and Looman mention is "a person with Static-99 score of 6 combined with a Psychopathy Checklist-Revised (PCL-R; Hare, 1993) score of 30." Some of the same data would contribute to a PCL-R score of 30 and to a Static-99 score of 6, so those two scores are not independent. It may well be that "most evaluat-ors agree" that a combination of scores such as that "warrants greater attention" and "causes more concern" than some other combinations. However, could an evaluator testify with a reasonable degree of certainty that a Static-99R risk assessment should be enhanced due to a score on the other test (PCL-R or plethysmograph)? We believe that an evaluator who chose to do so would have an obligation to explain (a) what incremental validity is, (b) the fact that there are insufficient data to calculate the incre-mental validity for such a combination, either for groups or individuals, and (c) that the opinion offered by the expert is just the expert's opinion, at best an "educated guess."

10. Wilson and Looman write (pp. 314-315), "Again, using Florida's SVP program as an example: Among a sample of 121 treatment participators sequentially assessed for treatment track placement, average Static-99 score was 5.96, aver-age score on the PCL-R was 22.72, and average Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 2000) was 27.61. This represents a high risk/needs population and, as such, we would be generally inclined to suggest that this is the normative sample to which these men should be compared."

Wilson and Looman include, in their Footnote 2: "We concede that the Courts have yet to fully accept all recent changes to the Static-99 protocol. Indeed, the Judge in a recent New Hampshire Daubert hearing (New Hampshire v. Thomas Hurley, 07-E-0236) accepted the new scorings based on age, as well as the new

recidivism estimates for 'routine' offenders (thus, acknowledging that reoffense rates have decreased). She did not, however, admit evidence regarding use of normative samples other than routine, citing '(un)reliable principles and methods' in making assignments. Rectifying this situation will require addition empirical validation of the process, specifically regarding clear delineation of inclusion criteria for each normative group and the establishment of inter-rater reliability coefficients with respect to the assignment process."

Our reading of these two passages is that Wilson and Looman are "generally inclined to suggest" that people who have been civilly committed as sexually violent predators would likely have a base rate of sexual reoffending more like the "High-Risk, High-Need" samples than the routine samples, but that additional empirical data might be needed before one could offer such an opinion in court. From our perspective, the important part is what can an evaluator say in court. We will briefly consider three stages of court proceedings.

First, consider an evaluator testifying at a trial in which a state is seeking to civilly commit a person as a sexually violent predator. Here, the proposition that civilly committed persons are more dangerous than routine sex offenders would be irrelevant to the risk assessment of this person, because he has not been civilly committed.

Second, consider an evaluator testifying at a hearing or trial in which a committed sexually violent predator who has not completed sex-offender treatment is seeking release. At this stage, the proposition that civilly committed persons are more dangerous than routine sex offenders would be relevant, if there is sufficient empirical support for the proposition.

Third, consider an evaluator testifying at a hearing or trial in which a committed sexually violent predator has completed the long-term sex-offender treatment program and is seeking release. At this stage, the proposition that civilly committed persons are more dangerous than routine sex offenders would be somewhat relevant, but a more important issue would be the detected sexual reoffense rate of people who have been civilly committed and have gone on to complete long-term sex-offender treatment.

11. As a final thought, Wilson and Looman write, "We are concerned that readers of the Campbell and DeClue paper might be unintentionally left with the impression that the original Static-99 scoring and norms, using a cutoff score of '6' projecting a 52 percent chance of reoffending in 15 years, represent the most accurate means of judging who will or will not sexually reoffend upon release to the community, at least using Static-99."

We agree that the original Static-99 norms are obsolete.

General Discussion

To clarify: We do not expect that an SVP evaluee's score on the Static-99 (or related instruments) can directly lead to a probability statement of his likelihood to sexually re-offend. We agree with the Static-99 Coding Rules (page 3; see www.static99.org) that "This instrument provides [the potential for] explicit probability estimates of sexual reconviction" for research samples. In our article critiqued by Wilson and Looman, and in this response to their critique, we explore how to maximize probability estimates from Static-99 and related instruments for the subjects in the research samples. We restate our recommendation regarding risk communication (from page 189 of our prior article):

> We do not believe that a forensic clinician must *insist* that the judge or jury make a decision that maximizes overall accuracy. We do not believe that a forensic clinician must *attempt to persuade* the judge or jury to make a decision that maximizes overall accuracy. We believe that a forensic clinician using an actuarial instrument in a risk assessment should *inform* the judge or jury how to maximize the accuracy of risk predictions using the risk-assessment tool. It is up to the judge or jury to make the final decision in a case. An expert conducting an actuarial risk assessment should communicate the findings in a way that shows how to maximize accuracy. Additional information, including tradeoffs associated with different cut scores, may be of interest to the decision maker, and experts should be prepared to describe the tradeoffs associated with using a cut score that does not optimize overall accuracy.

One of us (GD) has presented relevant tables and text from our "Maximizing Predictive Accuracy" article at two jury trials. The experience in the courtroom, questions posed by the jury to the expert at the end of testimony, and the juries' decisions all suggest that jurors are neither confused nor overwhelmed by frequency distributions and contingency tables presented at SVP trials.[11]

Received September 24, 2010; accepted September 29, 2010

---

[11] Incidentally, GD was called as a witness by the State in one case, and by the Respondent in the other case. In both of those cases, the jury decided to civilly commit the Respondent.

References

Andrews, D. A., & Bonta, J. (2000). *LSI-R: Level of Service Inventory—Revised.* Toronto, ON: Multi-Health Systems.

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.

Craig, L. A., Browne, K. D., & Stringer, I. (2004). Comparing sex offender risk assessment measures on a UK sample. *International Journal of Offender Therapy and Comparative Criminology, 48,* 7-27.

Hanson, R. K. (1997). The development of a brief actuarial risk scale for sexual offense recidivism. Ottawa, ON: Ministry of Public Safety and Emergency Preparedness 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., & Howard, P. D. (2010). Individual confidence intervals do not inform decision-makers about the accuracy of risk assessment evaluations. *Law and Human Behavior, 34*, 275-281.

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

Hare, R. (2003). *The Psychopathy Checklist—Revised* (2nd Edition). Toronto, ON: Multi-Health.

Hart, S. J., Webster, C. D., & Menzies, R. J. (1993). A note on portraying the accuracy of violence predictions. *Law and Human Behavior, 17,* 695-700.

Serin, R. C., & Brown, S. L. (2000). The clinical use of the Hare Psychopathy Checklist-Revised in contemporary risk assessment. In G. C. Gacono (Ed.), *The clinical and forensic assessment of psychopathy.* Mahwah, NJ: Lawrence Erlbaum Associates.

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