Maximizing Predictive Accuracy in Sexually Violent Predator Evaluations[1]

Authors: Terence W. Campbell & Gregory DeClue

Terence W. Campbell, Ph.D., ABPP (forensic), Sterling Heights, Michigan[2]
tcampbell3920@comcast.net

Gregory DeClue, Ph.D., ABPP (forensic), Sarasota, Florida

Abstract: The Static-99 and related procedures (Static-2002, Static-2002R, and Static-99R) are the most frequently used actuarial instruments for assessing sexual-recidivism risk. We analyze the original 15-year recidivism data for the Static-99 and all of the subsequent "norms" included in datasets from www.static99.org[3] and published articles. We use frequency tables to construct traditional 2 X 2 contingency tables for the datasets. We report and analyze corrected values for the Static-2002 dataset that was incorrectly presented in Hanson, Helmus, and Thornton (2010). We use traditional test utilities to seek optimal cutoff scores in order to maximize overall accuracy. We provide summary tables showing optimal cutoff scores for all four instruments. We provide an example that illustrates how an evaluator could use traditional methods for classification and prediction tests to report positive predictive value (PPV), with associated confidence intervals, as recommended by Heilbrun, Douglas, and Yasuhara (2009).

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

---

In discussing sexual-recidivism risk assessment in 1998, Hanson wrote (pp. 52-53), "All risk assessment should be informed by research, but the way to best apply research depends on the nature of the research available. . . . Given the current state of knowledge, I believe that there are three plausible approaches to conducting risk assessments: guided clinical,[4] pure actuarial, and adjusted actuarial." Our understanding is that research available now, including a recent meta-analysis (Hanson and Morton-Bourgon, 2009), supports the following propositions regarding the state of knowledge in 2010.

---

[1] Associate Editor's Note: The second author of this manuscript is the journal's editor and publisher. I supervised peer review, which was conducted by independent reviewers who were "blind" to the authorship of the article.

[2] Thanks from both authors to the editors and reviewers, who provided valuable criticism and guidance.

[3] Editor's Note: At the time of publication, this website was inactive. It is anticipated that the website will be reactivated. In the meantime, readers can access many of the documents from other websites using an Internet search engine and entering the title and/or a quoted section of each document.

[4] This is also called structured professional judgment.

Campbell, T. W., & DeClue, G. (2010). Maximizing predictive accuracy in sexually violent predator evaluations. *Open Access Journal of Forensic Psychology, 2,* 148-232.

In published, peer-reviewed research,[5] overall, on average,

1. The instruments in the Static-99 group[6] are about as accurate as, or more accurate than, other existing actuarial instruments.[7]

2. There is no evidence that adding an additional actuarial instrument improves accuracy, compared to one of these instruments alone (Seto, 2005).

3. There is no evidence that structured professional judgment (SPJ) is more accurate than the use of an instrument from the Static-99 group.[8]

4. There is no evidence that clinical adjustment to one of these instruments improves accuracy (Campbell & DeClue, 2010).

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. To put that in perspective, imagine that in your state 1,000 convicted sex offenders will be released this year. If the State wants to identify a subset of those sex offenders that is most likely to re-offend within the next five years, there is currently no procedure known to be more accurate than the instruments in the Static-99 group.

In this article, we explore the accuracy of the instruments in the Static-99 group. In doing so, we encounter the current limits of predictive power regarding sexual-recidivism risk assessment.

[5] Thornton, Hanson, & Helmus (n.d., p.2) mention "One of the symposia at the 2009 ATSA conference systematically examined the incremental predictive value of psychological risk factors, assessed either through questionnaires or ratings, relative to static actuarial assessment." We look forward to examining those data as articles are written, peer reviewed, and published.
[6] Static-99, Static-2002, Static-2002R, and Static-99R.
[7] However, in some studies, the Static-99 was not the most accurate predictor of sexual recidivism, or was not the most accurate for certain groups or subgroups. See, for example, Rettenberger, Matthes, Boer, & Eher (2009); Eher, Rettenberger, Schilling, & Pfäfflin (2008).
[8] SPJ was more accurate in a few studies. See, for example, de Vogel, de Ruiter, van Beck, & Mead (2004).

Actuarial Instruments

A. Relative Risk versus Absolute Risk

The instruments in the Static-99 group (see www.static99.org)[9] are the most frequently used and most widely researched actuarial instruments for assessing sex-offender recidivism risk (Helmus, Hanson, & Thornton, 2009). The Static-99 includes 10 items assessing static, historical factors (such as offense history, marital history, and victim gender). The predictive accuracy of the Static-99 was originally expressed in terms of relative risk (Hanson & Thornton, 2000). Relative risk refers to an offender's risk level compared to other offenders. Correlation coefficients, Area-under-the-Curve values (AUC), mean differences (Cohen's d), and regression coefficients have been used to identify the accuracy of relative-risk estimates (Helmus, 2009; Helmus, Hanson, & Thornton, 2009).

Unlike relative risk, absolute risk refers to the expected probability of sexual reoffending (e.g., 52% of offenders sharing this offender's offense history sexually reoffend over a 15-year follow-up). Actuarial risk instruments such as the Static-99 rely on explicitly defined rules identifying which risk factors are considered, how they are scored, and how the factors are combined mathematically to yield an objective risk estimate (Murrie, et al., 2009). In her master's thesis (which heralded the development of the Static-99R), Helmus (2009) noted, "Absolute risk information is required in certain high-stakes evaluations, notably sex offender civil commitment statutes in the United States" (p. 7-8).

In the United States, sex-offender civil-commitment statutes have been established in 20 states[10] and by the federal government. Typically, such civil commitment requires a finding that the person has a mental abnormality or personality disorder that makes him or her likely to engage in sexual violence if he or she is not confined. A person committed under such a law is labeled a "sexually violent predator" (SVP). The ultimate decision for the judge or jury is whether the person meets the jurisdiction's SVP criteria, and it is a binary decision: meets criteria or does not meet criteria. The risk assessment undergirding one prong of the ultimate decision is itself binary. Is this person considered likely[11] to engage in future acts of sexual violence if not confined: yes or no?

B. Binary Classification

In 1991, Birnbaum and Sheps presented a clear, concise introduction for professionals in their article, "Validation of New Tests." We recommend their article, and we use their ideas here. More recent descriptions are provided elsewhere (including encyclopedia

[9] Helmus, Hanson, and Thornton (2009) declare that www.static99.org is the "Static-99 official website."
[10] 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.
[11] "Likely" can be defined by statute or case law, or it can be deliberately left vague. In this article, unless otherwise stated, we treat "likely" as meaning "more likely than not."

articles on the Internet). For more sophisticated readers, we recommend Frederick and Bowden (2009). Our goal here is to provide an introduction that should be readily understood by clinical evaluators.

Birnbaum and Sheps (1991, pp. 622-625) write as follows:

> Tests may provide valuable information for identification and prediction functions. . . . The importance of information from specific tests depends on their validity and utility. . . . Poorly validated tests can produce harmful, misleading results. . . . A criterion standard, providing the so-called 'gold standard' cases, traditionally forms the basis for evaluating validity of diagnostic tests. This standard may be gold in name only because it is nothing more than the most accurate currently accepted test. Two x two tables are used to generate statistical summaries of accuracy from the perspective of a test (sensitivity and specificity) or from the perspective of test subjects and clinicians (predictive value of positive and negative results). Predictive values change as a function of sensitivity, specificity, and case prevalence [base rate]. . . . Precision of these estimates is a function of sample size, and confidence intervals are more informative than point estimates alone. Moreover, if cases or noncases selected for validation test sets are not typical of settings in which a test is to be used, various biases and limits to generalization may render the validation results specious. . . . It is important to remember that tests developed from mathematical models are forced to fit the data in training sets, so criterion validity should be assessed by applying the model to additional, independent data. . . . In conclusion, tests must be shown to be sufficiently precise, accurate, and reliable for their intended purpose.

An actuarial risk-assessment tool such as the Static-99 can be used in SVP cases to assist with the binary decision of whether the person is likely to engage in new acts of sexual violence if not confined. For each person facing SVP civil-commitment proceedings, one question is whether available evidence shows that he or she is likely to engage in sexual violence. Because there is no empirical evidence that adjusting Static-99 scores increases overall accuracy, we will explore the accuracy of the Static-99 when it is used as a pure-actuarial risk-assessment tool for SVP assessment. In this context, the Static-99 is a binary classification tool (Vrieze & Grove, 2008).

Binary classification is the task of classifying a group of people (here, people facing SVP commitment) into two groups (likely or not likely to engage in new acts of sexual violence) based on whether they have a particular property (a threshold score on the Static-99). As a classificatory or predictive instrument, the Static-99 relies on cut scores. All Static-99 scores above the cut score (e.g., 6 and above) rule in recidivism risk. All Static-99 scores below the cut score (e.g., 5 and below) rule out recidivism risk. Considerations of True Positive Rate (TPR) and False Positive Rate (FPR) influence the selection of cut scores. Cut scores can be selected to maximize TPR, to minimize FPR, or to maximize overall accuracy (Frederick & Bowden, 2009).

In this article, unless otherwise stated, we will focus on cut scores that maximize overall accuracy. We use detected sexual recidivism[12] as the criterion standard for sexual recidivism.

C. Exploring the Accuracy of Actuarial Instruments

When using an actuarial instrument, four outcomes can occur:

- True Positive: We predict the offender will reoffend, and he does reoffend.
- True Negative: We predict the offender will not reoffend, and he does not.
- False Positive: We predict the offender will reoffend, but he does not.
- False Negative: We predict the offender will not reoffend, but he does.

These values are typically presented in a 2 X 2 table such as Table 1.[13]

Table 1
A Model 2 X 2 Table for a Binary Classification Test

	Detected to Have Reoffended	Not Detected to Have Reoffended	
Predicted to Reoffend	**True Positive (TP)**	**False Positive (FP)**	→ Positive Predictive Value (PPV)
Not Predicted to Reoffend	**False Negative (FN)**	**True Negative (TN)**	→ Negative Predictive Value (NPV)
	↓ True Positive Rate (TPR)	↓ False Positive Rate (FPR)	

For a particular test, research data are plugged into the center four cells (TP, FP, FN, and TN). Then standard formulas can be used to calculate the following, all of which help users understand how accurate the test is at predicting recidivism.

- sample base rate (BR)
- true positive rate (TPR), also called sensitivity
- false positive rate (FPR)
- specificity
- positive predictive value (PPV)

[12] Generally, "detected sexual recidivism" is based on charges or convictions. This varies from study to study. See the original research articles for more specific descriptions of the samples.
[13] See, for example, Clinical Calculator 1 at http://faculty.vassar.edu/lowry/VassarStats.html

- negative predictive value (NPV)

The sample *base rate (BR)* is the proportion of people with a condition. Here, it is the number of people detected to sexually recidivate in a particular sample within a specified time frame.

The *true positive rate (TPR),* also called *sensitivity,* is the proportion of actual positives that are identified. Here, it is the proportion of sexual recidivists that were correctly predicted to sexually reoffend. It is calculated as TP / (TP + FN).

The *false positive rate (FPR)* is the proportion of people who were predicted to sexually reoffend, but who did not sexually reoffend. It is calculated as FP / (FP + TN).

Specificity is the proportion of people who did not sexually recidivate who were predicted not to sexually recidivate. It is calculated as FP / (FP + TN).

The *positive predictive value (PPV)* is the probability that a person will sexually reoffend, given a positive score on the test. It is calculated as TP / (TP + FP).

The *negative predictive value (NPV)* is the probability that a person will not sexually reoffend, given a negative score on the test. It is calculated as TN / (FN + TN).

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." Consequently, the general acceptance of 2 X 2 contingency tables for assessing classification accuracy when predicting violence in general, and sexual violence in particular, is beyond dispute. These tables identify the levels of predictive accuracy obtained for any given cutoff[14] score.

The prevailing paradigm for binary classification tests includes several assumptions, including the following (Frederick and Bowden, 2009):

- FPR and TPR are independent of any sample to which an individual belongs; therefore
- FPR and TPR are independent of local base rate (BR).
- However, overall accuracy is dependent on local base rate.

[14] Here, we use "cutoff score" and "cut score" interchangeably.

With those assumptions, researchers can evaluate the accuracy of different risk-assessment instruments. Clinical evaluators can use available research about the accuracy of various risk-assessment tools as they consider which tool to use in an applied context (e.g., the next person referred for a risk assessment).

D. Interpreting an Individual's Test Score

Heilbrun et al. (2009, p. 352) write the following:

> The appropriate use of actuarial strategies in a forensic context in which a prediction is requested involves using tools that have been derived and validated on appropriately large samples, and communicating in terms of probability or frequencies, while concurrently citing the margin of error, emphasizing the usefulness of extreme categories, and describing the increased uncertainty when group-based data are applied to single cases. If evaluators report precise probability estimates, they should also report the known confidence intervals and the implications of the confidence intervals for the accuracy of the estimate. If there are no known confidence intervals, however, then evaluators should not communicate a precise numerical estimate of risk because the precision of such an estimate cannot be evaluated.

After a particular risk-assessment instrument has been selected and an individual person's score is calculated, a utility derived from the classification table can help a clinician interpret the test score. Positive Predictive Value (PPV) tells the probability that the person would recidivate, given a positive test score. Negative Predictive Power (NPV) tells the probability that the person would not recidivate, given a negative test score. Confidence intervals help to communicate how precise (or imprecise) the probability estimate is.

The prevailing paradigm for binary classification tests includes these assumptions (Frederick and Bowden, 2009):

- PPV and NPV are dependent on the sample to which an individual belongs, therefore
- clinicians must know (or reasonably estimate) local BR in order to use a risk-assessment tool to guide risk assessment in any individual case.

We will consider individual test-score interpretation more fully after further exploration of classification accuracy.

The Original Static-99 Normative Data

A. Classification Tables

The original Static-99 normative data (Hanson & Thornton, 2000) consisted of 1,086 previously convicted sex offenders, followed for an average of 15 years. In this sample, 52% percent were from Canada, 48% were from the United Kingdom; 69% were child molesters, 31% were rapists. Recidivism was defined by conviction for this sample except for 142 Canadian offenders. For them, recidivism was defined by criminal charges or readmission to the facility. Table 2 summarizes the normative data for the Static-99 as reported by Hanson and Thornton (2000). In particular, this table summarizes the frequency of detected sex-offense recidivists and non-recidivists at each score. The data found in Table 2 allow computing the accuracy levels of the Static-99 using 2 X 2 contingency tables, across any cutoff score ranging from 1 to 6.

Table 2
Fifteen-Year Detected Sexual Recidivism Rate in the
Static-99 Normative Sample (Hanson & Thornton, 2000)

Score	N	15-year Detected Recidivism Rate	Detected to Have Reoffended	Not Detected to Have Reoffended
0	107	.13	14	93
1	150	.07	11	139
2	204	.16	33	171
3	206	.19	39	167
4	190	.36	68	122
5	100	.40	40	60
6+	129	.52	67	62
Totals	1,086	.25	272	814

Table 2 (above) shows that, for every Static-99 score below 6, most people are not detected to sexually reoffend. Slightly more than half of those with a score of 6 or higher were detected to sexually re-offend. If we decided to use a cutoff score of 6 on the Static-99 (predicting that everyone with a score of 6 or higher would re-offend and everyone with a score of 5 or lower would not re-offend), how many false positives, false negatives, etc., would occur?

Table 3 (below) helps us consider the accuracy of the Static-99 with a cutoff score of 6.

Table 3
Accuracy Levels for Static-99 Cutoff of 6 or Higher
Normative Sample, 15-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	67 (TP)	62 (FP)	129
Not Predicted to Reoffend	205 (FN)	752 (TN)	957
Totals	272	814	1086

Sample BR = .25, TPR (Sensitivity) = .25, FPR = .08, Specificity = .92
Overall Accuracy at this BR = .7541, PPV at this BR = .52, NPV at this BR = .79

With Table 3, we can see that about 52% (67 of 129) of the people predicted to reoffend were detected to reoffend. In other words, there are 67 true positives (TPs), slightly more than the 62 false positives (FPs). The table provides a way of addressing other important questions. Of the 272 people detected to reoffend, how many would be classified as predicted to reoffend based on their Static-99 scores? That is the true positive rate (TPR), also called Sensitivity. In this case, TPR is about .25 (67/272). Most (205 of 272, or about 75%) of the people who are detected to recidivate would be "missed" by the Static-99 if we use a cut score of 6. When we use a cut score of 6 to predict that someone will not sexually recidivate, we are correct 92% of the time (752 of 814). The false positive rate (FPR) is .08 (62/814). Specificity is .92 (752/814).

Overall, how accurate is classification using a cut score of 6 on the Static-99 when the base rate is .25? With 67 true positives and 752 true negatives, we correctly classify 819 of the 1086 cases, which is 75%. How does that compare to "betting the base rate"? Table 2 reminds us that if we simply predicted that no one would be detected to sexually reoffend, we would be correct for 814 of the 1086 cases. The difference in overall accuracy is trivial; either way (using a cut score of 6 or predicting that no one will be detected to recidivate) leads to about 75% overall accuracy.[15]

Positive Predictive Value (PPV) tells us the conditional probability that a person predicted to sexually recidivate will actually be detected to have sexually recidivated. This value varies with the base rate. PPV is the number of true positives (TP) divided by the total number predicted to recidivate (TP + FP). For the data in Table 3, PPV is 67/129, or .52.

Negative Predictive Value (NPV) tells us the conditional probability that a person predicted to not sexually recidivate will not be detected to have sexually recidivated. It, too,

[15] We recognize, of course, that there are tremendous practical implications to the choice of using a cut score of 6 versus predicting no one will recidivate, but at this point, we restrict our focus to classification accuracy.

varies with the base rate. NPV is the number of true negatives (TN) divided by the total number predicted to not recidivate (FN + TN). For the data in Table 3, NPV is 752/957, or .79.

What would happen if we wanted to decrease the number of false negatives (people predicted not to recidivate, but who are detected to recidivate)? Table 4 helps us see the accuracy tradeoffs that would ensue.

Table 4
Accuracy Levels for Static-99 Cutoff of 4 or Higher
Normative Sample, 15-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	175	244	419
Not Predicted to Reoffend	97	570	667
Totals	272	814	1086

Sample BR = .25, TPR (Sensitivity) = .64, FPR = .30, Specificity = .70
Overall Accuracy at this BR = .6860, PPV at this BR = .42, NPV at this BR = .85

What happens if we choose a cut score of 13 or higher for this sample? Because the highest score possible on the Static-99 is 12, a cut score of 13 means that we predict that no one will sexually recidivate.

Table 5
Accuracy Levels for Static-99 Cutoff of 13 or Higher
Normative Sample, 15-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	272	814	1086
Totals	272	814	1086

Sample BR = .25, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .7495, PPV is undefined, NPV at this BR = .75

Table 5 shows that, with a base rate of .25 and a cut score of 13, we would correctly classify 75% of the cases. With these data, “betting the base rate” is about as accurate as using a cut score of 6 (see Table 3). Using a cut score of 6 is slightly more accurate than predicting that no one will be detected to recidivate. Either way, overall accuracy is about 75%.

We calculated accuracy tables for other cut scores for the Static-99 normative samples (Hanson & Thornton, 2000). No other cut score in any of the samples (5-year, 10-year, 15-year) yielded greater overall accuracy than that obtained by using a cut score of either 6 or 13 in the 15-year sample (as shown in Tables 3 and 5, above).

B. Discussion

The data and analyses so far tell us that, with a base rate of detected sexual recidivism of .25, the overall accuracy of the Static-99 is, at best, 75%, which happens to be about equal to that of "betting the base rate," *i.e.,* estimating that no one will sexually re-offend. In samples with a higher base rate, the Static-99 can classify individuals with greater accuracy. In samples with a lower base rate, use of the Static-99 will inevitably lead to more false-positive classifications than true positives.

Recent Research with the Static-99

A. Lower Sexual-Recidivism Rates

In recent years, detected sexual recidivism rates have dropped dramatically. Helmus, Hanson, and Thornton (2009) reported that, for samples of sex offenders released from confinement more recently (1990 and later), detected sexual recidivism rates may be as much as 59% lower than those in the Static-99 original samples. One place to see varying sexual recidivism rates across samples is in Table 6 (below), which is adapted from Helmus's 2009 master's thesis (available at www.static99.org). *The reader should note that several of the datasets in the following table are from unpublished manuscripts (see References).*

Table 6
5-Year Recidivism Rates for Several Recent Studies
(Sorted from Lowest Detected Sexual Recidivism to Highest)
[Adapted From Helmus (2009) Appendix E]

Study	N	Detected to Have Recidivated (%)	Not Detected to Have Recidivated
Cortoni & Nunes (2007)	17	0 (0.0)	17
Hanson et al. (2007)	31	0 (0.0)	31
Eher et al. (2008)	151	3 (2.0)	148
Boer (2003)	299	11 (3.7)	288
Långström (2004)	1278	69 (5.4)	1209
Johansen (2007)	272	16 (5.9)	256
Ternowski (2004)	247	16 (6.5)	231
Craissati et al. (2008)	200	15 (7.5)	185
Swinburne Romine et al. (2008)	569	48 (8.4)	521
Bigras (2007)	207	19 (9.2)	188
Harkins & Beech (2007)	197	19 (9.6)	178
Epperson (2003)	150	16 (10.7)	134
Hill et al. (2008)	73	8 (11.0)	65
Allan et al. (2007)	298	35 (11.7)	263
Wilson et al. (2007a & b)	103	12 (11.7)	91
Bartosh et al. (2007)	90	12 (13.3)	78
Brouillette-Alarie & Proulx (2008)	199	29 (14.6)	170
Bonta & Yessine (2005)	101	19 (18.8)	82
Bengtson (2008)	310	61 (19.7)	249
Haag (2005)	198	39 (19.7)	159
Nicholaichuk (2001)	168	38 (22.6)	130
Knight & Thornton (2007)	433	107 (24.7)	326
22 studies	5,591	592 (10.6)	4,999

Range	Number of Studies
0-5%	4
5.01-10%	7
10.01-15%	6
15.01-20%	3
20.01-25%	2

In those studies, the 5-year detected sexual recidivism rate varied from 0 to 24.7%, with an overall recidivism rate of 10.6%.

B. Helmus (2009)

Helmus (2009, p. 32) reports data from 9,261 previously convicted sex offenders:

> Twelve samples were from Canada, six were from the United States, four were from the United Kingdom, and there was one each from Denmark, Austria, Holland, Sweden, Switzerland, Germany and New Zealand. Most offenders were from correctional settings (k=22), while 7 samples included offenders from mental health settings or mixed mental health and correctional settings. Of the 17 studies that that could be classified in terms of their treatment status, 9 samples were mostly treated (more than 75% of the sample), whereas 6 were mixed in their treatment exposure, and only 1 sample was mostly untreated.

The mean age for that sample was 40 (SD=12). Approximately 81 percent of this sample was released in 1990 or later. The average Static-99 score for this sample was 3.1 (SD=2.2).

Helmus (2009) shows sexual recidivism estimates from survival analysis for the original Static-99 development sample 1,086 people) and "current samples" (8,726 people) over 5, 10, and 15 years.[16] We encourage readers to examine that data (Helmus's Table 9, on page 59 of her 2009 master's thesis, posted at www.static99.org). In our Table 7, we present data from the 15-year follow-up of the 8,726 offenders in Helmus's "current samples."

Table 7
Sexual Recidivism Estimates from Survival Analysis at 15 Years
Adapted from Helmus (2009, p. 59)

Score	Initial N	Sexual Recidivism Rate	Sexual Recidivists	Not Sexual Recidivists
0	1,075	.055	59	1,016
1	1,375	.104	143	1,232
2	1,461	.086	126	1,335
3	1,386	.186	258	1,128
4	1,196	.204	244	952
5	842	.290	244	598
6+	1,391	.369	513	878
Totals	8,726	.18	1587	7,139

We constructed 2 X 2 contingency tables for the seven possible cut scores (0 through 6+). We present three of those tables in Appendix B.

[16] Helmus (2009) used survival analysis and Cox regression to compute estimates. See her thesis for details.

For this dataset, overall accuracy is maximized with a cut score of 13, as shown in Table B-1 (see Appendix B for tables beginning with "B"). No one would be predicted to sexually reoffend, and overall accuracy is 82%.

For this dataset, the next-best accuracy is obtained with a cut score of 6, as in Table B-2. With a base rate of .18 and a cut score of 6, overall accuracy is 78%. Of the 1,391 people who would be predicted to sexually reoffend, 513 (37%) of them were detected sexual recidivists. Using detected sexual recidivism as the criterion standard and a cut score of 6, predictions that a person in this sample would sexually recidivate would be wrong 63% of the time.

C. Detailed Recidivism Tables Static-99 (October 2008)

1. Combined Sample, 10-Year Follow-up

Posted on the Official Website of the Static-99, in a document called "Detailed Recidivism Tables Static-99 (October 2008)," Static-99 scores and recidivism data are presented for 1,621 offenders followed for 10 years. These data are labeled the "Complete Sample." The raw data for this sample are presented in Table 8.

Table 8
Static-99 Scores and Observed Sexual Recidivism within 10 Years (Complete Sample)
Developed from Data Reported in Detailed Recidivism Tables Static-99 (October 2008)

Score	N	Detected Recidivism Rate	Detected to Have Sexually Recidivated	Not Detected to Have Sexually Recidivated
0	175	.06	10	165
1	196	.05	9	187
2	226	.07	15	211
3	229	.17	40	189
4	272	.18	49	223
5	178	.29	51	127
6	150	.33	49	101
7	105	.45	47	58
8	60	.37	22	38
9	16	.56	9	7
10	14	.50	7	7
Totals	1,621	.19	308	1313

Earlier, we explored the accuracy of the Static-99 at various cut scores for the original sample, which had a detected sexual recidivism rate of .25. Now we explore the accuracy of the Static-99 when the detected sexual recidivism rate is .19. How accurate is the Static-99 when the base rate (BR) is .19, with a cutoff score of 6? See Table 9 (below).

Table 9
Accuracy Levels for Static-99 Cutoff of 6 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
Complete Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	134	211	345
Not Predicted to Reoffend	174	1,102	1,276
Totals	308	1,313	1,621

Sample BR = .19, TPR (Sensitivity) = .44, FPR = .16, Specificity = .84
Overall Accuracy at this BR = .7625, PPV at this BR = .39, NPV at this BR= .86

The sample base rate (BR) is .19, meaning that 19% of the people in this sample were detected to sexually reoffend within the 10-year follow-up period. An evaluator who predicted that none of these people would be detected to sexually re-offend within the 10-year period would be correct 81% of the time. Could use of the Static-99 with a cut score of 6 improve the accuracy of prediction? As summarized in the table above, the true positive rate (TPR) is .44 (of the 308 people detected to have sexually reoffended, the Static-99 correctly identifies 44% of them). The false positive rate (FPR) is .16 (of the 1,313 people who were not detected to have sexually reoffended, 16% would be incorrectly predicted to sexually reoffend). The overall accuracy (true positives plus true negatives; 134 + 1,102) is .76, which is a decrease in overall accuracy compared to relying on the base rate alone. With a cut score of 6, the positive predictive value is 134/345, or .39. Of the 345 people who would be predicted to sexually reoffend, 134 (39%) of them were detected to have sexually reoffended. Using detected sexual recidivism as our criterion standard, the positive predictive value (PPV) tells us how many predictions that someone would sexually recidivate would be correct: 39%.[17]

We calculated test utilities for the other possible cut scores (any score from 0 to 10) for this sample. Tables B-4 through B-7 (in Appendix B) present the findings for cut scores of 4, 8, 9, and 13 respectively.

For the 1,621 people included in the Complete Sample, 19% of them were detected to sexually recidivate within 10 years. Using detected sexual recidivism as our criterion standard, overall accuracy with the Static-99 is maximized with a cut score of 9. (This is slightly more accurate than predicting that no one will sexually re-offend.) With a cut score of 9, 30 (5%) of the people in the sample would be classified as predicted to sexually reoffend, and 16 (52%) of those predictions would be accurate.

[17] We encourage readers new to this type of analysis, and those who have not used these skills in recent years, to make sure that they understand the terms in this example before moving on. In the remainder of this paper, we assume that readers are familiar with terms and abbreviations such as BR, TPR, PPV, etc.

Discussion

During the first decade following its development and dissemination, the Static-99 became the most widely used instrument for assisting with sexual-recidivism risk assessment. As mentioned in the introduction, there has been no general finding that any other tool is more accurate. There has been no consistent empirical finding that the accuracy of predictions based on the Static-99 is improved by adding another risk-assessment instrument or by adjusting the risk estimate via clinical judgment.

Recent research studies have reported lower overall detected sexual recidivism rates (base rates), compared to those in the Static-99 development sample. The above 2 X 2 contingency tables and related test utilities illustrate that, as with any prediction tests, the farther the base rate is from .50, the less accurate the Static-99 is for estimating absolute risk.

Another aspect of recent sexual-recidivism research is that, beyond the fact that the base rate of detected sexual recidivism is lower in recent studies, it is also more variable. We mentioned this briefly when we presented Table 6 (above), adapted from Helmus (2009).

The Official Static-99 Website (www.static99.org) provides a potentially great service to evaluators and others by sorting their 2008 samples and reporting findings at different base rates. They provide frequency tables for samples with sexual-recidivism base rates varying from 7% (5-Year Routine Correctional-Services-of-Canada [CSC] Samples) to 29.8 (10-Year High-Risk Samples). Our thanks to those at the website for sorting and presenting the data should not be construed as endorsement of their recommendations for interpreting the data.

Next, we explore two datasets that are subsamples drawn from the Combined Sample mentioned above.

2. Routine CSC Sample, 10-Year Fixed Follow-up

Posted on the Official Website of the Static-99, in a document called “Detailed Recidivism Tables Static-99 (October 2008),” Static-99 scores and recidivism data are presented for 342 offenders followed for 10 years. These data are labeled the “Routine CSC Sample.” The raw data for this sample are presented in Table 10.

Table 10
Static-99 Scores and Observed Sexual Recidivism within 10 Years
Detailed Recidivism Tables Static-99 (October 2008)
Routine CSC Sample

Score	N	Detected Recidivism Rate	Detected to Have Sexually Recidivated	Not Detected to Have Sexually Recidivated
0	42	.048	2	40
1	54	.019	1	53
2	49	.020	1	48
3	45	.089	4	41
4	56	.071	4	52
5	34	.029	1	33
6	32	.125	4	28
7	17	.529	9	8
8	10	.200	2	8
9	1	.000	0	1
10	2	.500	1	1
Totals	342	.085	29	313

Of these 342 cases, 29 (8.5%) were detected to sexually reoffend within 10 years. Tables B-8 through B-11 (in Appendix B) are contingency tables for cut scores of 6, 7, 8, or 13.

For the 342 people included in the 2008 Static-99 Routine CSC Sample with 10-year fixed follow-up, 8.5% of them were detected to sexually recidivate within 10 years. Using detected sexual recidivism as our criterion standard, overall accuracy with the Static-99 is maximized with a cut score of 13. No one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 91.5% of those predictions are accurate.

The next most accurate classification comes with a cut score of 7. Thirty (9%) of the people in the sample would be classified as predicted to sexually reoffend, and 12 (40%) of those predictions would be accurate. With a cut score of 7, overall accuracy (90% correct) would be lower than predicting that no one would be detected to sexually recidivate within 10 years (91.5% correct).

Additional research with larger samples might yield different results.

3. High-Risk Sample, 10-Year Fixed Follow-up

Posted on the Official Website of the Static-99, in a document called “Detailed Recidivism Tables Static-99 (October 2008),” Static-99 scores and recidivism data are presented for 735 offenders followed for 10 years. These data are labeled the “High-Risk

Sample." The detected sexual recidivism rate is .298. The raw data for this sample are presented in Table 11.

Table 11
Static-99 Scores and Observed Sexual Recidivism within 10 Years
Detailed Recidivism Tables Static-99 (October 2008)
High-Risk Sample

Score	N	Detected Recidivism Rate	Detected to Have Sexually Recidivated	Not Detected to Have Sexually Recidivated
0	50	.120	6	44
1	38	.079	3	35
2	71	.169	12	59
3	92	.250	23	69
4	154	.273	42	112
5	110	.391	43	67
6	93	.366	34	59
7	66	.424	28	38
8	37	.432	16	21
9	15	.600	9	6
10	9	.333	3	6
Totals	735	.298	219	516

Of these 735 cases, 219 (30%) were detected to sexually reoffend within 10 years. Tables B-12 through B-16 (in Appendix B) are contingency tables for cut scores of 7, 8, 9, 10, or 13.

For the 735 people included in the 2008 Static-99 High-Risk Sample with 10-year fixed follow-up, 29.8% of them were detected to sexually recidivate within 10 years. Overall accuracy with the Static-99 is maximized with a cut score of either 9 or 13. With a cut score of 9, 24 (3%) of the people in the sample would be classified as predicted to sexually reoffend, and 12 (50%) of those predictions would be correct. With a cut score of 13, no one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard and using a cut score of either 9 or 13, 70% of predictions are accurate.

Additional research with larger samples might yield different results.

4. General Discussion about the Static-99

Helmus, Hanson, & Thornton (2009) offer recommendations for how evaluators could report Static-99 in light of recent research. We consider Helmus, Hanson, and Thornton to be experts on the development and scoring of the Static-99. We certainly recommend that anyone using the Static-99 should be guided by them, along with others who developed the instrument and the coding rules, for guidance on scoring the Static-99.

With rare exceptions (e.g. clerical errors, overlooking of a detail), test users should generally follow the test developers' rules when administering and scoring tests.

We consider test interpretation to be less "owned" by the test developers. For example, Helmus et al. (2009, p. 38) note, "Crime rates peaked in the early 1990s and have been generally declining since then." They note some possible explanations, and their guesses are as good as ours (perhaps yours are better). If we knew why crime rates, including sex-crime rates, have declined, we might be better at interpreting Static-99 scores in samples with different base rates.

Helmus et al. (2009, pp. 40-41) write,

> Differences in recidivism within each Static-99 score on the basis of sample type and offender type suggest that evaluators can no longer, in an unqualified way, associate a single Static-99 score with a single recidivism estimate. Instead, each Static-99 score is associated with a range of recidivism estimates, and evaluators must make a separate judgment as to where a particular offender lies within that range. This new conceptualization of recidivism norms forces evaluators to consider factors external to the risk scale. Although the best method of considering these external factors is as yet unknown, there are several factors worth considering in this decision. These factors include the risk-relevant characteristics of the population from which the offender is selected (as described above), as well as risk-relevant characteristics of individual offenders.

We do not concur with that advice. Consistent with decades of work with classification and prediction tests by researchers in medicine, psychology, and other disciplines, we recommend that evaluators who use the Static-99 use traditional classification tables and test utilities to guide their risk assessments and risk communications. Our analyses reveal that the overall accuracy of risk predictions based on Static-99 scores varies across samples with different base rates, *just as it should for any classification or prediction test.* There is nothing new about that, and a "new conceptualization of recidivism norms" is not warranted. Rather than speculate about unknown external factors (the dark matter of sexual-recidivism risk assessment?), 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.

Static-2002

Static-2002 is a 14-item actuarial instrument. Scores range from 0 to 14. According to the Coding Rules for Static-2002 on the www.static99.org website (pages 1-4; accessed April 9, 2010):

> Static-2002 is an actuarial risk tool for evaluating the risk of sexual and violent recidivism among adult male sexual offenders (Hanson & Thornton, 2003). Like Static-99, Static-2002 can be used by a wide range of evaluators (e.g., psychologists, probation officers, psychiatrists, therapists) using commonly available criminal history information. Static-2002 predicts sexual, violent, and any recidivism as well as other actuarial risk tools commonly used with sexual offenders (Hanson & Morton-Bourgon, 2009) and slightly better than Static-99 (Hanson, Helmus, & Thornton, 2010; Helmus, 2007). . . . To better understand what is being measured, Static-2002 items are grouped into five domains: Age, Persistence of Sex Offending, Deviant Sexual Interests, Relationship to Victims, and General Criminality.

Hanson et al. (2010) have reported some Static-2002 data for a sample of 3,034 previously convicted sex offenders. The average age at release for this sample was 39 (SD=12). For 65% of the sample (n=1,985), recidivism was defined as new charges. Convictions (n=1,049) served as the recidivism criterion for the remaining 35%. Table 12 shows detected-recidivism data for 1,923 people followed up for 5 years.[18]

[18] This is a corrected table. The corresponding data reported in Hanson et al. (2010) are incorrect. Personal communication from Dawne E. Amsler and Leslie Helmus, 4/15/2010.

Table 12

Static-2002 Scores and Observed Sexual Recidivism within 5 Years
Adapted from Hanson et al.'s (2010) Table 6

Score	N	Detected Recidivism Rate	Detected to Have Sexually Recidivated	Not Detected to Have Sexually Recidivated
0	30	.033	1	29
1	106	.019	2	104
2	176	.045	8	168
3	193	.052	10	183
4	247	.101	25	222
5	282	.145	41	241
6	271	.133	36	235
7	206	.175	36	170
8	200	.245	49	151
9	99	.333	33	66
10	70	.386	27	43
11	31	.258	8	23
12	11	.545	6	5
13	1	.000	0	1
Totals	1,923	.145	282	1,641

Of these 1,923 cases, 282 (14.7%) were detected to sexually reoffend within 5 years. Tables B-17 through B-20 (in Appendix B) are contingency tables for cut scores of 8, 10, 12, or 15.

For the 1,923 people included in the and Hanson et al. (2010) sample with 5-year fixed follow-up, 14.7% of them were detected to sexually recidivate within 10 years. Overall accuracy with the Static-2002 is maximized with a cut score of either 12 or 15. With a cut score of 12, 12 (<1%) of the people in the sample would be classified as predicted to sexually reoffend, and 6 (50%) of those predictions would be correct. With a cut score of 15, no one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 85% of predictions are accurate.

Additional research with larger samples might yield different results.

Static-2002R

A document titled "Reporting Static-2002R scores with 2009 recidivism norms (non-routine): A template for cases for which the norms for routine samples do not apply" (Phenix, Helmus, Hanson, & Thornton, 2009a) includes the following:

> The Static-2002R is an instrument designed to assist in the prediction of sexual and violent recidivism for sex offenders. Given that Static-2002R was found to fully incorporate the relationship between age at release and

> sexual recidivism, whereas the original Static-2002 scale did not (Thornton, Helmus, & Hanson, 2009), the developers of Static-2002 recommend that the revised version of the scale (Static-2002R) replace Static-2002 in all contexts where it is used. . . . Hanson and Thornton (2003) developed this risk assessment instrument based on follow-up studies from Canada, the United States, and the United Kingdom with a total sample size of 2,169 sexual offenders from 10 samples. Using seven replication samples (n = 2,605), Static-2002R demonstrated moderate to large accuracy in the prediction of sexual, violent, and general recidivism. The Static-2002R consists of 14 items and produces estimates of relative risk based upon the number of risk factors present in any one individual. The risk factors included in the risk assessment instrument are grouped into five domains: age, persistence of sex offending, deviant sexual interests, relationship to victims, and general criminality.

Possible scores range from -2 to 13. At the time of this writing (April 2010), www.static99.org includes several Static-2002R frequency tables. Here (A, B, and C, below), we analyze three of them.

A. Five-Year Sexual Recidivism Rates for Static-2002R: Routine Sample

A document titled, "Reporting Static-2002R scores with 2009 recidivism norms (routine samples): A template for cases for which the norms for routine samples apply" (Phenix, Helmus, Hanson, & Thornton, 2009b) includes the following:

> This group consisted of three samples of sex offenders from Canada. 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 pre-selection processes. The recidivism norms for the unselected samples are the closest available to a hypothetical average of all offenders.

Observed sexual recidivism rates are reported for 526 people followed for 5 years.[19]

[19] At the time of this writing (April 2010), no data are provided at www.static99.org for longer follow-up for the Routine Sample.

Table 13
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

Of these 526 cases, 28 (5.3%) were detected to sexually reoffend within 5 years. Tables B-21 through B-24 (in Appendix B) are contingency tables for cut scores of 7, 9, 11, or 14.

For the 526 people included in the Routine Sample with 5-year fixed follow-up, 5.3% of them were detected to sexually recidivate within 5 years. Overall accuracy with the Static-2002 is maximized with a cut score of either 11 or 14. With a cut score of 11, 2 (<1%) of the people in the sample would be classified as predicted to sexually reoffend, and 1 (50%) of those predictions would be correct. With a cut score of 14, no on would be predicted to sexually reoffend, and 95% of those predictions would be accurate.

B. Ten-Year Sexual Recidivism Rates for Static-2002R: Non-routine Sample

A document titled “Reporting Static-2002R scores with 2009 recidivism norms (non-routine): A template for cases for which the norms for routine samples do not apply” (Phenix et al., 2009a) includes the following:

> The non-routine group includes all samples of offenders preselected in some way (4 samples). It therefore combines samples preselected based on treatment need, as well as those preselected as high risk/high need. In some cases there may have been some measure of preselection and the

offender would be most similar to either the preselection for treatment or pre-selection for high-risk/needs samples. If the amount of preselection is unknown and there is no strong evidence to differentiate between preselected for treatment and preselected for high-risk/needs then the non-routine sample norms are an option to consider.

Observed sexual recidivism rates are reported for 766 people followed for 10 years.

Table 14
Static-2002R Scores and Observed Sexual Recidivism within 10 Years
Non-routine Sample

Score	N	Detected Recidivism Rate	Detected to Have Sexually Recidivated	Not Detected to Have Sexually Recidivated
-2	1	.000	0	1
-1	4	.025	1	3
0	21	.095	2	19
1	45	.044	2	43
2	62	.113	7	55
3	77	.104	8	69
4	121	.298	36	85
5	101	.317	32	69
6	95	.253	24	71
7	70	.329	23	47
8	81	.407	33	48
9	39	.385	15	24
10	32	.656	21	11
11	11	.182	2	9
12	5	.600	3	2
13	1	.000	0	1
Totals	766	.273	209	557

Of these 766 cases, 209 (27.3%) were detected to sexually reoffend within 10 years. Tables B-25 through B-28 are contingency tables for cut scores of 8, 10, 12, or 14.

For the 766 people included in the Static-2002R Non-routine Sample with 10-year fixed follow-up, 209 (27.3%) of them were detected to sexually recidivate within 10 years. Overall accuracy is maximized with a cut score of 10. With a cut score of 10, 49 (6%) of the people in the sample would be classified as predicted to sexually reoffend, and 26 (53%) of those predictions would be correct.

The next most accurate classification (slightly less accurate than a cut score of 10) comes with a cut score of 12 or 14. With a cut score of 12, 6 (1%) of the people in the sample would be classified as predicted to sexually reoffend, and 3 (50%) of those predictions would be correct. With a cut score of 14, no one is predicted to sexually reof-

fend. With detected sexual recidivism as the criterion standard, 73% of those predictions are accurate.

Additional research with larger samples might yield different results.

C. Ten-Year Sexual Recidivism Rates for Static-2002R: High-Risk/Need Sample

A document titled "Reporting Static-2002R scores with 2009 recidivism norms (non-routine): A template for cases for which the norms for routine samples do not apply" (Phenix et al., 2009a) includes the following:

> This would include a small minority of offenders selected from routine correctional populations on the basis of risk and need factors external to Static-2002R. Offenders in this group were referred for services at forensic psychiatric facilities such as offenders referred as Mentally Disorder Sex Offenders, Sexually Violent Predators/Sexually Dangerous Persons, Not Guilty by Reason of Insanity, or for treatment of a mental disorder (sexual or otherwise). It would also include offenders referred to intensive treatment programs reserved for the highest risk offenders (not moderate intensity treatment programs, or treatment programs offered to the majority of sex offenders). Offenders identified as high risk through a quasi-judicial or administrative process examining a range of risk relevant characteristics such as sentence extensions for dangerousness (e.g., preventative or indefinite detention, treatment orders, denial of statutory release) would also be included in this group.

Observed sexual recidivism rates are reported for 642 people followed for 10 years.

Table 15
Static-2002R Scores and Observed Sexual Recidivism within 10 Years
High-Risk/Need Sample

Score	N	Detected Recidivism Rate	Detected to Have Sexually Recidivated	Not Detected to Have Sexually Recidivated
-2	0	.000	0	0
-1	0	.000	0	0
0	18	.111	2	16
1	24	.042	1	23
2	44	.159	7	37
3	62	.129	8	54
4	106	.311	33	73
5	84	.345	29	55
6	88	.250	22	66
7	63	.333	21	42
8	72	.389	28	44
9	34	.353	12	22
10	31	.677	21	10
11	10	.200	2	8
12	5	.600	3	2
13	1	.000	0	1
Totals	642	.294	189	453

Of these 642 cases, 189 (29.4%) were detected to sexually reoffend within 10 years. Tables B-29 through B-32 (in Appendix B) are contingency tables for cut scores of 8, 10, 12, or 14.

For the 642 people included in the Static-2002R High-Risk/Need Sample with 10-year fixed follow-up, 189 (29.4%) of them were detected to sexually recidivate within 10 years. Overall accuracy is maximized with a cut score of 10. With a cut score of 10, 47 (7%) of the people in the sample would be classified as predicted to sexually reoffend, and 26 (55%) of those predictions would be correct.

The next most accurate classification (slightly less accurate than a cut score of 10) comes with a cut score of 12 or 14. With a cut score of 12, 6 (1%) of the people in the sample would be classified as predicted to sexually reoffend, and 3 (50%) of those predictions would be correct. With a cut score of 14, no one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 71% of those predictions are accurate.

Static-99R

Beginning in late 2009, the developers of the Static-99 recommended that clinicians switch to the newly developed Static-99R. The items on the Static-99R are identical to those on the original Static-99, except for Item 1. On item 1 of the Static-99, the person gets either 0 or 1 point, with 1 point going to those younger than age 25. For the Static-99R, a person can get a score of 1, 0, -1, or -3, depending on his age. Total scores for the Static-99R can range from -3 to +12. Otherwise, the scoring of the Static-99R items is identical to the scoring of the original Static-99 items. (See the Coding Rules at www.static99.org for additional information.)

The following description is posted at www.static99.org in "Static-99R Reporting Template: Routine Samples (Word or PDF)":

> Static-99R . . . is an actuarial measure of relative risk for sexual offense recidivism. Given that Static-99R was found to fully incorporate the relationship between age at release and sexual recidivism, whereas the original Static-99 scale did not (Helmus, 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. Static-99R has shown moderate accuracy in ranking offenders according to their relative risk for sexual recidivism. . . . There have been a large number of studies examining the absolute sexual recidivism rates associated with Static-99 scores. Helmus (2009) combined 28 Static-99 replication studies and was able to calculate Static-99R scores for 23 of these samples. The samples (*n* = 8,139) were drawn from Canada, the United States, United Kingdom, western Europe and New Zealand. Recidivism was defined as charges in about half of these studies and as convictions in the other half. . . . Although the relative risk was consistent across studies, the observed recidivism base rates varied considerably across samples based on factors not measured by Static-99R. Samples that were preselected to be high-risk/high needs (6 samples) show the highest recidivism rates, samples preselected based on treatment need (6 samples) had intermediate recidivism rates, and routine correctional samples had recidivism rates substantially lower than the preselected groups (and also lower than the recidivism rates in the original development samples for Static-99). . . . In applying the recidivism norms it is ideal to use local norms that are applicable to the group of offenders to which this offender most closely resembles. Given that these norms are not often available, the routine sample will usually reflect the most appropriate recidivism rates as they are representative of typical sex offenders in correctional systems.

Along with new rules for scoring Item 1 come new recommendations for interpreting a person's total score. Hansen, Phenix, and Helmus (2009, pp. 19-20) addressed "What's an Evaluator to Do?" They suggest that evaluators can "focus on relative risk"

and that "any statements about absolute risk requires (sic) justification."[20] They describe three options for statements about absolute risk:

> Option 1: Use local "norms."
> Option 2: Use routine "norms."
> Option 3: Justify that routine "norms" do not apply.[21]

Option 1: Use local "norms."

Hansen et al. (2009, p. 19) consider the use of local norms to be "ideal, but not often possible."

Option 2: Use routine "norms."

Hansen et al. (2009, p. 20) write, "The estimates for routine samples are the default position. [They are] representative of general population of adjudicated sex offenders. This option is sufficient in most circumstances."

Option 3: Justify that routine "norms" do not apply.

Hansen et al. (2009) suggest three possible "justifications":

- Sufficient criminogenic needs to recommend treatment: use treatment need norms
- Member of small minority selected on risk/need factors external to Static-99R: use high risk/need norms
- Sufficient evidence that offender is non-routine, but insufficient information to differentiate between treatment need or high risk/need: use non-routine norm

In the absence of local "norms," an evaluator who is required to address absolute risk can use the routine "norms" (the default position) or use other "norms" if there is justification to do so. In deciding which "norms" to use, the evaluator should consider the various samples that comprised the four groups of samples:

1. Routine Corrections
2. Non-Routine
3. Preselected Treatment Need
4. Preselected High-Risk/Need[22]

[20] Page 19.
[21] Pages 19-20 of Static-99(R) and Static-2002(R): How to Interpret and Report in Light of Recent Research (October 2009) at http://www.static99.org/.
[22] Page 12 of Static-99R Evaluators Workbook at http://www.static99.org/.

A. Routine Sample

The following description is posted at www.static99.org in "Static-99R Reporting Template: Non-routine Samples (Word or PDF)":

> 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.

The "Norms" tab of the Official Website of the Static-99 includes a seven-page pdf, Detailed recidivism tables Static-99R (October 2009). For the Routine Sample, only five-year data are included. Table 16 shows the frequency distribution for the 2,406 people in the fixed follow-up sample.

Table 16
Static-99R Scores and Observed Sexual Recidivism within 5 Years
Detailed Recidivism Tables Static-99R (October 2009)
Routine Sample

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

Of these 2,406 cases, 145 (6%) were detected to sexually reoffend within 5 years. Tables B-33 through B-36 (in Appendix B) are contingency tables for cut scores of 6, 9, 10, or 13.

For the 2,261 people included in the 2009 Static-99R Routine Sample with 5-year fixed follow-up, 145 (6%) of them were detected to sexually recidivate within 5 years. Overall accuracy with the Static-99R is maximized with a cut score of 13. No one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 94% of those predictions are accurate.

The next most accurate classification (slightly less accurate than a cut score of 13) comes with a cut score of 10. With a cut score of 10, 8 (<1%) of the people in the sample would be classified as predicted to sexually reoffend, and 2 (25%) of those predictions would be correct.

Additional research with larger samples might yield different results.

B. Preselected-for-Treatment Sample

The following description is posted at www.static99.org in "Static-99R Reporting Template: Non-routine Samples (Word or PDF)":

> This group consisted of six samples of offenders referred for sex offender specific treatment during their current incarceration. If an offender is selected for treatment but does not receive it due to bed shortages, he would still be considered preselected for treatment. It is the selection that defines this sample, not the participation in treatment. This includes referral for community sex offender treatment programs for any type of conditional release during the current incarceration or for non-custodial sentences. The quality of the treatment program, jurisdiction of the program, program structure (length or content), and the quality of the offender's participation in and completion of the program is not a consideration in the definition of this group. These factors would be taken into account by an evaluator outside of the Static-99R assessment. This sample is defined by the presence of treatment needs. Samples were categorized in this group if the treatment program was specific to sex offenders and offenders were referred for treatment during their current incarceration. Given the overlap in dynamic risk factors between sex offenders and general offenders, it is plausible that offenders referred to other (i.e., non-sex-offender-specific) treatment programs may be similar to this group. Additionally, offenders referred for treatment during previous incarcerations could also plausibly fit in this group given that at some point they were identified as having treatment needs warranting intervention and that they subsequently reoffended.

The "Norms" tab of the Official Website of the Static-99 includes a seven-page pdf, Detailed recidivism tables Static-99R (October 2009). For the Preselected-for-Treatment Sample, five-year and ten-year data are reported separately. Table 17 shows the frequency distribution for the 866 people in the ten-year, fixed follow-up sample.

Table 17
Static-99R Scores and Observed Sexual Recidivism within 10 Years
Detailed Recidivism Tables Static-99R (October 2009)
Preselected-for-Treatment Sample

Score	N	Detected Recidivism Rate	Detected to Have Sexually Recidivated	Not Detected to Have Sexually Recidivated
-3	16	.000	0	16
-2	12	.167	2	10
-1	85	.047	4	81
0	114	.088	10	104
1	131	.069	9	122
2	142	.141	20	122
3	110	.118	13	97
4	111	.126	14	97
5	67	.284	19	48
6	37	.324	12	25
7	17	.412	7	10
8	18	.389	7	11
9	4	.250	1	3
10	1	.000	0	1
11	1	.000	0	1
Totals	866	.136	118	748

Of these 866 cases, 118 (13.6%) were detected to sexually reoffend within 10 years. Tables B-37 through B-41 are contingency tables for cut scores of 6, 8, 9, or 13.

For the 866 people included in the 2009 Static-99R Preselected-for-Treatment Sample with 10-year fixed follow-up, 118 (13.6%) of them were detected to sexually recidivate within 10 years. Overall accuracy with the Static-99R is maximized with a cut score of 13. No one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 86% of those predictions are accurate.

The next most accurate classification (slightly less accurate than a cut score of 13) comes with a cut score of 9. With a cut score of 9, 6 (1%) of the people in the sample would be classified as predicted to sexually reoffend, and 2 (33%) of those predictions would be correct.

Additional research with larger samples might yield different results.

C. High-Risk/Need Sample

The following description is posted at www.static99.org in "Static-99R Reporting Template: Non-routine Samples (Word or PDF)":

> This would include a small minority of offenders selected from routine correctional populations on the basis of risk and need factors external to Static-99R. Offenders in this group were referred for services at forensic psychiatric facilities, such as offenders referred as Mentally Disorder (sic) Sex Offenders, Sexually Violent Predators/Sexually Dangerous Persons, Not Guilty by Reason of Insanity, or for treatment of a mental disorder (sexual or otherwise). It would also include offenders referred to intensive treatment programs reserved for the highest risk offenders (not moderate intensity treatment programs, or treatment programs offered to the majority of sex offenders). Offenders identified as high risk through a quasi-judicial or administrative process examining a range of risk relevant characteristics such as sentence extensions for dangerousness (e.g., preventative or indefinite detention, treatment orders, denial of statutory release) would also be in this group.

The "Norms" tab of the Official Website of the Static-99 includes a seven-page pdf, Detailed recidivism tables Static-99R (October 2009). For the High-Risk/Need Sample, five-year and ten-year data are reported separately. Table 18 shows the frequency distribution for the 703 people in the ten-year, fixed follow-up sample.

Table 18
Static-99R Scores and Observed Sexual Recidivism within 10 Years
Detailed Recidivism Tables Static-99R (October 2009)
High-Risk/Need Sample

Score	N	Detected Recidivism Rate	Detected to Have Sexually Recidivated	Not Detected to Have Sexually Recidivated
-3	0	.000	0	0
-2	2	.000	0	2
-1	24	.083	2	22
0	31	.097	3	28
1	56	.089	5	51
2	42	.333	14	28
3	82	.268	22	60
4	130	.262	34	96
5	133	.293	39	94
6	83	.386	32	51
7	61	.475	29	32
8	32	.375	12	20
9	17	.471	8	9
10	9	.444	4	5
11	1	.000	0	1
Totals	703	.290	204	499

Of these 703 cases, 204 (29.0%) were detected to sexually reoffend within 10 years. Tables B-41 through B-44 are contingency tables for cut scores of 6, 9, 10, or 13.

For the 703 people included in the 2009 Static-99R High-Risk/Need Sample with 10-year fixed follow-up, 204 (29%) of them were detected to sexually recidivate within 10 years. Overall accuracy with the Static-99R is maximized with a cut score of 13. No one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 71% of those predictions are accurate.

The next most accurate classification (slightly less accurate than a cut score of 13) comes with a cut score of 10. With a cut score of 10, 10 (1%) of the people in the sample would be classified as predicted to sexually reoffend, and 4 (40%) of those predictions would be correct.

Additional research with larger samples might yield different results.

D. Non-Routine Sample

The following description is posted at www.static99.org in "Static-99R Reporting Template: Non-routine Samples (Word or PDF)":

> The non-routine group includes all samples of offenders preselected in some way. It therefore combines samples preselected based on treatment need, as well as those preselected as high risk/high need, and also includes a small number of offenders preselected in different ways that fit neither category (e.g., preselected based on offence severity). In some cases there may have been some measure of preselection and the offender would be most similar to either the preselection for treatment or pre-selection for high-risk/ high-needs samples. If the amount of preselection is unknown and there is no strong evidence to differentiate between preselected for treatment and pre-selected for high-risk/high-needs then the non-routine sample norms are an option to consider.

The "Norms" tab of the Official Website of the Static-99 includes a seven-page pdf, Detailed recidivism tables Static-99R (October 2009). For the Non-Routine Sample, five-year and ten-year data are reported separately. Table 19 shows the frequency distribution for the 1,626 people in the ten-year, fixed follow-up sample.

Table 19
Static-99R Scores and Observed Sexual Recidivism within 10 Years
Detailed Recidivism Tables Static-99R (October 2009)
Non-Routine Sample

Score	N	Detected Recidivism Rate	Detected to Have Sexually Recidivated	Not Detected to Have Sexually Recidivated
-3	16	.000	0	16
-2	14	.143	2	12
-1	109	.055	6	103
0	146	.089	13	133
1	187	.075	14	173
2	188	.181	34	154
3	199	.181	36	163
4	250	.196	49	201
5	214	.280	60	154
6	130	.369	48	82
7	86	.430	37	49
8	53	.377	20	33
9	22	.409	9	13
10	10	.400	4	6
11	2	.000	0	2
Totals	1,626	.204	332	1,294

Of these 1,626 cases, 332 (20.4%) were detected to sexually reoffend within 10 years. Tables B-45 through B-48 are contingency tables for cut scores of 6, 9, 10, or 13.

For the 1,626 people included in the 2009 Static-99R Non-Routine Sample with 10-year fixed follow-up, 332 (20.4%) of them were detected to sexually recidivate within 10 years. Overall accuracy with the Static-99R is maximized with a cut score of 13. No one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 80% of those predictions are accurate.

The next most accurate classification (slightly less accurate than a cut score of 13) comes with a cut score of 10. With a cut score of 10, 12 (1%) of the people in the sample would be classified as predicted to sexually reoffend, and 4 (33%) of those predictions would be correct.

Additional research with larger samples might yield different results.

Summary of Analyses

The next two tables summarize our analyses.

Table 20
Summary of Accuracy Levels for Static-99 and Static-2002

Instrument (Sample)	Fixed Follow-up (years)	N	BR	Optimum Cut Score to Maximize Overall Accuracy	Overall Accuracy at Optimum Cut Score	Percent of Sample Predicted to Sexually Recidivate	PPV	For More Info See Table
Static-99								
(Hanson & Thornton, 2000)	15	1,086	.250	6	.7541	12	.52	3
(Helmus, 2009, p. 59)	15	8,726	.182	13	.8181	0	NA	B-3
(October 2008, Combined Sample)	10	1,621	.190	9	.8112	5	.53	B-6
(October 2008, Routine CSC)	10	342	.085	13	.9152	0	NA	B-11
(October 2008, High-Risk)	10	735	.298	9[23]	.7020	3	.50	B-14
(October 2008, High-Risk)	10	735	.298	13	.7020	0	NA	B-16
Static-2002								
(Hanson et al., 2010)	5	1,923	.147	12[24]	.8534	<1	.50	B-19
(Hanson et al., 2010)	5	1,923	.147	15	.8534	0	NA	B-20

N = number of subjects, BR = sample base rate, PPV = positive predictive value

[23] Accuracy is maximized with a cut score of either 9 or 13. We present data for each of those cut scores here.
[24] Accuracy is maximized with a cut score of either 12 or 15. We present data for each of those cut scores here.

Table 21
Summary of Accuracy Levels for Static-2002R and Static-99R

Instrument (Sample)	Fixed Follow-up (years)	N	BR	Optimum Cut Score to Maximize Overall Accuracy	Overall Accuracy at Optimum Cut Score	Percent of Sample Predicted to Sexually Recidivate	PPV	For More Info See Table
Static-2002R								
(October 2009, Routine)	5	526	.053	11[25]	.9468	<1	.50	B-23
(October 2009, Routine)	5	526	.053	14	.9468	0	NA	B-24
(October 2009, Routine)	10	766	.273	10	.7311	6	.53	B-26
(October 2009, Non-routine)	10	642	.294	10	.7134	7	.55	B-30
Static-99R								
(October 2009, Routine)	5	2,406	.060	13	.9397	0	NA	B-36
(October 2009, Preselected for Treatment)	10	866	.136	13	.8637	0	NA	B-40
(October 2009, High-Risk/Need)	10	703	.290	13	.7098	0	NA	B-44
(October 2009, Non-routine)	10	1,626	.204	13	.7958	0	NA	B-48

N = number of subjects, BR = sample base rate, PPV = positive predictive value

[25] Accuracy is maximized with a cut score of either 11 or 14. We present data for each of those cut scores here.

We continue to focus on maximizing overall accuracy in this section, using detected sexual recidivism as the criterion standard.

With the original sample of the Static-99, 12% of the sample would be classified as likely to sexually reoffend, with a positive predictive value (PPV) of .52. In other words, based on actuarial assessment alone, 12% of the people would be considered likely – a bit more likely than not – to sexually reoffend within 15 years.

For these instruments in the more recent (2008 to 2009) samples, 0% to 7% of the people would be classified as likely to sexually reoffend, and only about half of those classifications would be correct. The greatest number of people would be classified as predicted to sexually reoffend when using the Static-2002R for the High Risk/Need sample. Seven percent (47 of 642) of the people would be classified as predicted to sexually reoffend, but only four percent of the entire sample (26 of 642) would be correctly classified as predicted to sexually reoffend.

Compared to the original Static-99 samples, the overall accuracy of the instruments is enhanced by using higher cutoff scores. Higher cutoff scores decrease the frequency of false positive predictions and increase the frequency of true negative predictions. We will consider the effect on public safety in the general discussion section, which follows.

General Discussion

A. Importance of Accuracy

Classification and prediction are important in many disciplines, including medicine, psychology, law, political science, and economics. Across fields, accuracy matters. In his testimony before the bipartisan Financial Crisis Inquiry Commission, Alan Greenspan, former chairman of the Federal Reserve (Fed) commented that, with hindsight, "I was right 70 percent of the time, but I was wrong 30 percent of the time." Commission member Brooksley Born, the former head of the Commodity Futures Trading Commission, said Greenspan's stewardship of the Fed had led the US into financial meltdown:

> The Fed utterly failed to prevent the financial crisis. The Fed and the banking regulators failed to prevent the housing bubble. They failed to prevent the predatory lending scandal. They failed to prevent our biggest banks and bank-holding companies from engaging in activities that would bring them to the verge of collapse without massive taxpayer bailouts. They failed to recognize the systemic risk posed by an unregulated over-the-counter derivatives market, and they permitted the financial system and the economy to reach the brink of disaster.[26]

Closer to home, Janus and Meehl (1997, p. 34) highlight the importance of accuracy in risk predictions in SVP cases:

[26] http://www.democracynow.org/2010/4/8/headlines#6

> There is a continuing debate about whether predictions of dangerousness are accurate enough to support deprivation of liberty. As in other civil commitment settings, the stakes in these determinations are high. The debate is especially important in the context of sex offender commitments, because the consequences of the predictions are so severe. If predictions about future violence are too optimistic, sexual violence may result. Unduly pessimistic predictions result in unnecessary, prolonged deprivations of liberty. In addition, sex offender commitments entail treatment that is expensive and intrusive, while sexual violence is exceedingly destructive. Thus, both types of prediction errors are costly in many ways.

Throughout this paper, we strive to present data and analyses that are understandable and uncontroversial. As we move into a general discussion, we have the same aims.

In an important recent paper, Thornton, Hanson, and Helmus (n.d., p. 1) discuss the importance of recent (2008 and 2009) research regarding base rates:

> There were two key findings regarding base rates. First, the mean sexual recidivism base rate after 5-years and 10-years (sic) follow up was lower in a newly assembled collection of 28 Static-99/recidivism samples than it was in Hanson and Thornton (2000). Second, there was highly significant variation in sexual recidivism base rates among the 28 samples. A detailed account of the analyses of these data may be found in Helmus (2009) and specific normative recidivism estimates my be downloaded from

http://www.static99.org/pdfdocs/st-99rworkbookwithsamplesandsummaries.pdf

> These findings of variable base rates have profound implications for actuarial risk assessment. They mean that in order to obtain a recidivism estimate, in addition to knowing how someone scores on Static-99, you also need to know the base rate for the population from which the offender comes.

We completely and wholeheartedly agree. This is exactly what should be expected for any classification or prediction test. As was mentioned earlier, Birnbaum and Sheps explained this in 1991. They noted that some test utilities, including true-positive rate and false-positive rate, are primarily of interest to researchers or others who compare one test to another. Some of those test utilities are not expected to vary with changes in base rates. Other test utilities, including the ones that are of most importance to clinicians and test subjects, do vary with changes in base rate. Positive predictive value, negative predictive value, and overall accuracy of a test all vary along with changes in the base rate.[27]

[27] They also vary with changes in cutoff score, as was illustrated earlier.

If all groups of sex offenders were pretty much alike, then each group of sex offenders might have similar sexual-recidivism rates. That is not the case, and some groups of sex offenders have higher detected sexual-recidivism rates than others.

With varying base rates, the only rationale for expecting that a risk-assessment tool would not have different PPVs and other test utilities at different base rates would be that the risk-assessment tool included all relevant risk factors. That is, not just every risk factor that we understand and can reliably and conveniently measure, but every risk factor that there is. It should not be surprising that none of the Static-99 group contain every factor that affects whether or not a person will sexually reoffend, over time, if he has opportunities. Like other classification and prediction tests, the instruments in the Static-99 group are incomplete and imperfect.

As it is with other classification and prediction tests, the positive predictive value (PPV) of a particular score varies with the base rate. To recap part of the quote above: "These findings of variable base rates have profound implications for actuarial risk assessment. They mean that in order to obtain a recidivism estimate, in addition to knowing how someone scores on Static-99, you also need to know the base rate for the population from which the offender comes" (Thornton et al., n.d., p. 1).

B. Confidence

As we have discussed, the positive predictive value (PPV) in a prediction test tells the probability that an event will occur, given a positive test score. With the instruments in the Static-99 group, using detected sexual recidivism as our criterion standard, PPV tells the probability that a person in the sample will sexually recidivate, given a positive score on the instrument. How confident should an evaluator, judge, or jury be about the precision of that test score? We explore that through a hypothetical cross-examination (set in 2010) of an expert witness in a trial involving civil commitment as a sexually violent predator (SVP). See Appendix A.

C. Risk Communication

With only rare exceptions, the accuracy of the instruments in the Static-99 group do not exceed the accuracy obtained by relying on the applicable base rates alone. Although there is no dispute about that fact, there is some controversy about what it means and what people should do about it. Here, we briefly recap some of the controversy and offer our own views.

While challenging actuarial instruments such as the Static-99, Vrieze and Grove (2008, p. 275) write, "Comparing the CF [Correct Fractions] of an instrument to the CF of betting the base rate is, quite frankly, a not very demanding validity hurdle." For Vrieze and Grove (2008), demonstrating that an actuarial instrument outperforms the base rate establishes its "incremental validity." Vrieze and Grove (2008, p. 275) further explain, "If an instrument will not allow one to materially outperform the base rate . . ., then one's test really has not much going for it."

Mossman (2008) insists that Vrieze and Grove (2008) have misinterpreted the statutory schemes providing for civil commitment of previously convicted sex offenders. Mossman contends that SVP statutes seek to identify which offenders are likely to reoffend. Mossman clarifies his position, insisting that SVP legal proceedings are not obligated "to improve on base rates or decide how to balance false positive and false negative decisions" (p. 287) and "Vrieze and Grove go astray by imposing their own view about how prediction instruments for sex offender recidivism should be used, viz., to optimize the fractions of correct predictions. Whatever one thinks about such a policy, mental health professionals must recognize a different one; identifying individuals who are 'likely to reoffend'" (p. 288).

Citing Monahan and Walker (1994), Heuer, Penrod, and Kattan (2007, pp. 573-574) wrote the following:

> Courts, including the U. S. Supreme Court, have repeatedly reviewed cases in which individual rights are pitted against the state's authority to restrict those rights when a predictive tool suggests that the individual is committing or is likely to commit a criminal act. . . . The courts have typically described their review process as a balancing of several considerations: the risk of a false positive error; the harm to the target . . . caused by a false positive error; and the offsetting societal gain achieved by using the predictive technology for the purpose being pursued by the law. . . . What the courts are asked to decide in such cases is the propriety of a legal procedure, and they have described their own decision making in these cases as a utilitarian balancing of societal benefits and individual harms.

Throughout this article, we have sought to maximize overall accuracy of risk predictions using various actuarial tools. At this point, we want to clarify our position regarding risk assessment and risk communication. We believe that a forensic clinician who uses an actuarial tool to conduct a risk assessment should communicate the results in a way that shows the trier of fact how to maximize the accuracy of predictions using the instrument.

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.

We urge readers to take note of the following, from Heilbrun et al. (2009, p. 351):

> Actuarial strategies developed from small samples and insufficiently validated may have wide margins of error. This margin of error for a 95% confidence interval, which is closely related to the precision of the prediction and the confidence with which it can be communicated, is wider for individual cases and small groups than it is for larger groups. Predictions about a given individual using such actuarial strategies are less likely to characterize accurately his or her violence risk if the probability is cited but these two caveats are not included in the communication.

D. Effects on Public Safety

Some people may be concerned that, if overall accuracy is maximized, fewer convicted sex offenders would be civilly detained than was the case a few years ago (when the original Static-99 development sample was used to guide risk predictions). Would such fears be warranted?

Consider 1,000 sex offenders scheduled to be released from confinement in a U. S. state in the year 2000. This imagined state has an SVP law, and evaluators use the Static-99 to estimate sexual recidivism. Considering the data and analyses from Tables 2 and 3, we can see that, with a cut score of 6 on the Static-99, 119 of those 1,000 sex offenders would be predicted to sexually reoffend within 15 years.[28] Let us imagine that those 119 people—no more and no less—would be civilly committed.

Now consider 1,000 sex offenders scheduled to be released from confinement in the same U. S. state in the year 2010. The imagined state still has an SVP law. Now evaluators use the Static-2002R to estimate sexual recidivism, using the current (2009) base rate for the Routine Sample. Considering the data and analyses in Tables 13 and B-23, 4 people would be predicted to sexually reoffend within 5 years. Let us imagine that those 4 people – no more and no less – would be civilly committed.

In these scenarios, nearly 30 times fewer people would be civilly committed in 2010 than in 2000. How many more sex crimes should we expect? First, we make a few assumptions, all of which are reasonable, and any of which might be untrue. We will assume that each "detected" sexual recidivist (charged or convicted) is guilty of 1 new sex crime, and each person not detected to sexually recidivate is guilty of 0 new sex crimes. We will assume that everyone who is recommended for civil committed is civilly committed, and no one is released from that indefinite confinement. We assume that the estimated base rates for sexual reoffending are accurate. We note that an evaluator in 2000 could use 15-year recidivism data for a routine sex offender but, in 2010, the only available data for a routine sex offender are 5-year recidivism data.

[28] Table 3 shows us that 129 of 1,086 (11.9%) people would be predicted to sexually reoffend.

What happens in the 2000 scenario? The base rate of sexual re-offending was then estimated to be .25, so if no one had been civilly committed, 250 people would have sexually reoffended. Because of the SVP law, 119 were civilly committed, and 52% of them, or 62 people, were true positives (see Table 3 for test utilities). With our assumptions, 62 sex crimes would be prevented, but 188 (250-62) would still occur. We correctly confine 62 people and unnecessarily confine 57 people. We prevent 62 sex crimes. Of the 881 sex offenders released from prison and not civilly committed, 188 go on to commit new sex crimes.

What would happen in the 2010 scenario? The base rate of sexual re-offending is estimated to be .04, so if no one were civilly committed, 40 people would be expected to sexually reoffend. Based on the data in Tables 13 and B-23, 4 people would be civilly committed, 2 of whom would have sexually reoffended if not confined (true positives). With our assumptions, 2 sex crimes would be prevented, but 38 (40-2) would still occur. We correctly confine 2 people and unnecessarily confine 2 people. We prevent 2 sex crimes. Of the 996 sex offenders released from prison and not civilly committed, 38 go on to commit new sex crimes.[29]

Considering the 2000 and 2010 scenarios together, 119 people would be civilly committed in 2000, but just 4 people would be civilly committed in 2010. Nearly 30 times fewer people would be civilly committed in 2010 than in 2000. Nevertheless, society would be safer in the 2010 scenario than in the 2000 scenario, with only about $^1/_5$ as many new sex crimes committed by the released sex offenders (38 by the sex offenders released in 2010, compared to 188 by those released in 2000).

We suspect that this example overstates the point, primarily because the only available data for routine sexual offenders in 2010, with either the Static-2002R or the Static-99R, are the 5-year sexual-recidivism data. We expect that, when sufficient data are collected to compute 10-year and 15-year rates, the "modern" sexual recidivism rates will be greater than .05 or .06, but still considerably less that the .25 from the original Static-99 sample. We note that evaluators conducting risk assessments in 2010 are faced with this exact problem. The Static-99R and the Static-2002R are the instruments recommended by the test developers. However, for estimating sexual recidivism for routine sex offenders in 2010, there are no current, reliable data beyond these 5-year sexual-recidivism data.

The bottom line is that maximizing overall accuracy produces a win-win when the base rate of sexual recidivism decreases, as it has in recent years. Fewer people would be recommended for civil commitment, but that does not lead to an increased threat to public safety. If decisions are made that maximize overall accuracy, when base rates of sexual recidivism decrease, fewer people are civilly committed. Nevertheless, fewer sex crimes are committed, and society is safer.

[29] An astute reader may have noticed that, based on the data reported in Tables B-23, B-24, and 21, cut scores of 11 and 14 are equally accurate. We used a cut score of 11 in the example above. If a cut score of 14 were used, no one would be civilly committed, and we would expect 40 new sex crimes from the 1,000 released sex offenders.

Appendix A
Cross Examination

The following is a hypothetical cross-examination (set in 2010) of an expert witness in a trial involving civil commitment as a sexually violent predator (SVP).

Attorney: Doctor, on direct examination, you testified about your clinical opinion regarding Mr. X, and about an actuarial instrument. For now, I'd like to ask you questions that are just about your actuarial assessment.

Doctor: Okay.

Attorney: I've read your written report, Doctor. Could you tell us how you selected which actuarial instrument to use?

Doctor: The most widely used and most widely researched actuarial tools regarding sexual recidivism are the Static-99 and related tests. A current article in the journal *Law and Human Behavior* by Hanson, Helmus, and Thornton, reports that, in a meta-analysis involving eight samples, Static-2002 was more accurate at predicting sexual recidivism than Static-99. That would suggest that one should choose the Static-2002 over the Static-99.

Within the past year, the developers of both Static-99 and Static-2002 now recommend the revised instruments—Static-99R and Static-2002R—in all contexts where the instruments are used. So, I chose to use the Static-2002R.

Attorney: Doctor, are you familiar with research regarding combining actuarial instruments, and whether that increases accuracy?

Doctor: Yes. There was a study by Seto in 2005. He found that, once you select the best actuarial instrument for your population, adding an additional actuarial instrument does not increase accuracy.

Attorney: And so, you used one actuarial instrument, the Static-2002R?

Doctor: Yes. That's what I routinely use in my practice these days, and that's what I used regarding Mr. X.

Attorney: Doctor, the developers of the Static-2002R recommend that you use local norms, is that correct?

Doctor: Yes, whenever possible or practical.

Attorney: And in this case?

Doctor: Unfortunately, we don't have local norms in this area at this time.

Attorney: So, the next best thing --?

Doctor: Yes. Mr. X was in the general population at the state prison. I considered the available data about Mr. X, along with the guidelines at the Static-99 website, and I decided to use the Routine "Norms." Those are the ones that are recommended for most sex offenders, and I consider them to be the most appropriate for Mr. X.

Attorney: Okay. Now, for the routine norms, are there any frequency tables available for a fixed follow-up period, to show how many people with each Static-2002R score were actually detected to sexually reoffend within a certain time frame?

Doctor: Yes.

Attorney: For what time frames are frequency data available?

Doctor: Right now, just the five-year time frame.

Attorney: So, you have frequency data for a five-year, fixed follow-up. And does that mean that a group of people were released from prison, and somebody checked to see how many of them were re-arrested or re-convicted for a sexual charge?

Doctor: Basically, that's it.

Attorney: How many people were in that group that was followed for five years?

Doctor: It was 526 people.

Attorney: Now, we're here in a court proceeding in which the jury [or judge] will decide whether Mr. X meets criteria for civil commitment as a sexually violent predator. Part of that determination involves the question of whether he is likely to commit a new act of sexual violence if he is not confined. That's the part of the statute that you used this Static-2002R for, isn't that right, Doctor?

Doctor: Yes.

Attorney: Doctor, you're aware that in this context the Static-2002R can be considered to be a classification test, correct?

Doctor: Yes.

Attorney: Now, I'm aware that you used your clinical judgment in your final analysis in this case. But for now, we're just considering the Static-2002R that you considered to be the most appropriate risk-assessment tool for this case. Doctor, are you familiar with the term "cut score" or "cutoff score" in classification tests?

Doctor: Yes.

Attorney: Once you've chosen a risk-assessment tool and you've identified or estimated the base rate of sexual reoffending in the population, you can figure out the cutoff score that maximizes overall accuracy, isn't that right, Doctor?

Doctor: Yes.

Attorney: And by "maximizing overall accuracy," that just means that you find the cutoff score that leads to the most correct predictions in that particular sample or group, right?

Doctor: That's right.

Attorney: Fine. Now, what is the base rate of detected sexual reoffending in that routine sample of folks on the Static-2002R?

Doctor: Well, that's the five-year sexual-recidivism base rate. It's about .04.

Attorney: .04. That's 4%, right, meaning that for the relevant comparison group, 4% of those people were charged with or convicted of a new sexual offense within five years of their release from confinement?

Doctor: Yes, that's right.

Attorney: And so 96% were not? Ninety-six percent of those routine sex offenders did not sexually reoffend within 5 years?

Doctor: Well –

Attorney: Oh, excuse me. I'll rephrase that. Ninety-six percent of those routine sex offenders were not charged or convicted for a new sex offense within 5 years of their release from confinement?

Doctor: Yes, that's right.

Attorney: Okay, now back to the cutoff score. What's the cutoff score that maximizes overall accuracy for this sample that we're talking about, the score that leads to the most accurate predictions overall for this group?

Doctor: That's a cut score of 11.

Attorney: Doctor, did you find that Mr. X met that cut score?

Doctor: Yes, I did. His score on the Static-2002R is 11, so he meets the cut score.

Attorney: Okay, now, is there a test utility that tells you, for the sample studied, what is the probability that the person really will sexually recidivate, given the fact that he has a positive score on the test?

Doctor: Yes. That's the positive predictive value, what we call PPV.

Attorney: So, you've chosen the test, the Static-2002R, and you've estimated the base rate by choosing the routine sample. You're using the cutoff score that leads to the most accurate predictions. And now you can tell us the probability of someone in that sample actually sexually recidivating, given that we know he got a positive score—11—on the test. So what's the number, what's that probability?

Doctor: It's .50.

Attorney: So, based on this risk-assessment instrument – just the instrument for now, not your clinical judgment – it's 50-50 whether somebody with a score of 11 will sexually recidivate within 5 years?

Doctor: With detected sexual recidivism as the criterion standard, that's right.

Attorney: Now, I want to ask you about confidence in that prediction. For now, I'm not asking about confidence in terms of your subjective, clinical judgment. I'm asking about statistical confidence, what statisticians call "confidence intervals." Are you familiar with confidence intervals, Doctor?

Doctor: Yes, I am.

Attorney: Good. Doctor, with classification and prediction tests, is it generally recommended that clinicians report not just a single score, but also the confidence intervals[30]?

Doctor: Yes, that's right.

Attorney: Now, is it possible to calculate confidence intervals for PPV, the positive predictive value, that you testified is .50 in this case?

Doctor: Yes, it's possible.

Attorney: And that can actually be done fairly easily at one of the online statistical calculators, correct?

Doctor: That's right.

Attorney: Doctor, are you familiar with the online statistical calculators at VassarStats[31]?

[30] Birnbaum and Sheps (1991) recommend this on page 624.

Doctor: Yes, I am.

Attorney: Have you ever used those online statistical calculators at VassarStats in your work?

Doctor: Yes, I have.

Attorney: Have those calculators been recommended to you and other forensic psychologists at professional workshops?

Doctor: Yes.

Attorney: Doctor, I'm going to ask you about those confidence intervals in just a minute or two, but let's put things in perspective just a bit, first. Are classification and prediction tests used in psychology?

Doctor: Yes.

Attorney: And in other fields, too, for example, medicine?

Doctor: Yes.

Attorney: And other scientific and professional fields?

Doctor: Yes.

Attorney: Doctor, when there are two possible outcomes, such as "has the disease" or "doesn't have the disease," the classification test is commonly called a binary classification test, isn't that right?

Doctor: Yes.

Attorney: One binary decision could be "likely to sexually reoffend" or "not likely to sexually reoffend," right?

Doctor: Yes.

Attorney: Doctor, is it fair to say that you used an actuarial test, the Static-2002R, to help you—and this jury [judge]—with a binary decision: Is Mr. X likely, or not likely, to sexually reoffend?

Doctor: You could put it that way. Yes, that's fair to say.

---

[31] http://faculty.vassar.edu/lowry/clin1.html#return. See the "Technical Note on Calculation of Confidence Intervals" there.

Attorney: So when we talk about statistics for a binary classification test, terms like overall accuracy, positive predictive value, and confidence interval, those are not obscure, arcane terms from some tiny branch of science, are they?

Doctor: No.

Attorney: In fact, those are the same terms that scientists and professionals use to communicate about how accurate their tests are, their tools for classifying things and predicting events?

Doctor: Yes, that's right.

Attorney: And no matter what the field, those statistics are calculated in the same way. You count the hits and misses, the true positives and false negatives and so on, and construct a 2 X 2 table?

Doctor: Yes, that's right.

Attorney: And once you have that 2 X 2 table, you can calculate the test utilities, like overall accuracy and so on?

Doctor: Yes.

Attorney: Okay, now a very brief review, and then we'll get to those confidence intervals we've been waiting for. You evaluated Mr. X, including a clinical interview and a review of his records?

Doctor: Yes.

Attorney: You determined that you had enough information to score an actuarial test, you chose the Static-2002R, you used the base rate for the routine correctional sample, and you used the cutoff score that leads to the most accurate predictions?

Doctor: Yes.

Attorney: And you found that Mr. X meets the cutoff, which leads to the prediction that, just considering the actuarial results, the best estimate of his likelihood to sexually recidivate within the next 5 years is about 50-50, right?

Doctor: As I said before, using detected sexual recidivism as the criterion standard, that's correct.

Attorney: Very well. Now let's talk about those confidence intervals. First, the numbers. The 5-year, fixed follow-up for routine corrections with the Static-2002R, you get a 2 X 2 table, and from that you get the positive predictive value of .50 that says a person

in that sample with a score of 11 or higher on the Static-2002R has a 50% probability of sexually reoffending within 5 years, right?

Doctor: Yes.

Attorney: Doctor, I'm showing you a printout of VassarStats Clinical Calculator 1. Do you recognize that?

Doctor: Yes.

Attorney: That's filled in with numbers for True Positives, False Negatives, and so on. Can you verify that those are the correct numbers for the 5-Year Routine Sample for the Static-2002R?

Doctor: Yes, those are the right numbers.

Attorney: Okay, now from that VassarStats Clinical Calculator 1, is there a number corresponding to "For a particular positive result, the probability that it is a True Positive"?

Doctor: Yes.

Attorney: What is that number?

Doctor: It's 0.5.

Attorney: That's the same number you told us for PPV, right? Same answer you calculated?

Doctor: Yes.

Attorney: And that's another way of saying 50%, right, 50-50?

Doctor: Yes.

Attorney: Now, just to the right of that 0.5 for PPV are some confidence intervals, do you see those?

Doctor: Yes.

Attorney: What numbers does it show for the 95% confidence intervals?

Doctor: It shows 0.026677 to 0.973323.

Attorney: Could you please round those numbers off to the nearest hundredth?

Doctor: Okay. That would be .02 to .97.

Attorney: So, in other words, 2% to 97%?

Doctor: Yes, that's right.

Attorney: So, this is based on the actuarial test that you consider to be the best and most accurate for this case. A person with a score of 11, like Mr. X got on the test, would have about a 50-50 probability of sexually re-offending within 5 years – that's the best single estimate or guess, right?

Doctor: That's right.

Attorney: And with the confidence intervals we would say that, statistically, it's 95% percent sure that the PPV for a person with that score on that test would be .02 to .97?

Doctor: Yes.

Attorney: So, if we just take the best statistical evidence and make the best specific estimate, the probability is about 50-50, but if we want to be more precise, the probability that someone with that score would sexually reoffend could be anywhere from 2% to 97%?

Doctor: Yes.

Attorney: That's what we get straight from the statistics, no clinical judgment, no subjective guess or gut feelings, just the straight statistical evidence?

Doctor: Yes, you could say that.

Attorney: Anywhere from 2% to 97%?

Doctor: Yes.

Attorney: Thank you, Doctor. No further questions.

Appendix B
Contingency Tables

I. Static-99

A. Helmus (2009)

Table B-1
Accuracy Levels for Static-99 Cutoff of 4 or Higher
Helmus (2009) Dataset, 15-Year Follow-up

	Sexual Recidivists	Not Sexual Recidivists	Totals
Predicted to Reoffend	1,001	2428	3,429
Not Predicted to Reoffend	586	4,711	5,297
Totals	1,587	7,139	8,726

Sample BR = .18, TPR (Sensitivity) = .63, FPR = .34, Specificity = .66,
Overall Accuracy at this BR = .6546, PPV at this BR = .29, NPV at this BR = .89

As shown in Table B-1 (above), with a base rate of .18 and a cut score of 4, overall accuracy is 65%. Of the 3,429 people who would be predicted to sexually reoffend, 1,001 (29%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 71% of the time.

Table B-2
Accuracy Levels for Static-99 Cutoff of 6 or Higher
Helmus (2009) Dataset, 15-Year Follow-up

	Sexual Recidivists	Not Sexual Recidivists	Totals
Predicted to Reoffend	513	878	1,391
Not Predicted to Reoffend	1,074	6,261	7,335
Totals	1,587	7,139	8,726

Sample BR = .18, TPR (Sensitivity) = .32, FPR = .12, Specificity = .88
Overall Accuracy at this BR = .7763, PPV at this BR = .37, NPV at this BR = .85

As shown in Table B-2 (above), with a base rate of .18 and a cut score of 6, overall accuracy is 78%. Of the 1,391 people who would be predicted to sexually reoffend, 513 (37%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 63% of the time.

Table B-3
Accuracy Levels for Static-99 Cutoff of 13 or Higher
Helmus (2009) Dataset, 15-Year Follow-up

	Sexual Recidivists	Not Sexual Recidivists	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	1,587	7,139	8,726
Totals	1,587	7,139	8,726

Sample BR = .18, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .8181, PPV is undefined, NPV at this BR = .82

As shown in Table B-3 (above), with a base rate of .18 and a cut score of 13, overall accuracy is 82%. No one would be predicted to sexually reoffend. With detected sexual recidivism as our criterion standard, 82% of those predictions would be accurate.

B. Combined Sample (October 2008), 10-Year Follow-up

Table B-4
Accuracy Levels for Static-99 Cutoff of 4 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
Complete Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	234	561	795
Not Predicted to Reoffend	74	752	826
Totals	308	1,313	1,621

Sample BR = .19, TPR (Sensitivity) = .76, FPR = .43, Specificity = .57,
Overall Accuracy at this BR = 6083, PPV at this BR = .29, NPV at this BR = .91

As shown in Table B-4 (above), with a base rate of .19 and a cut score of 4, overall accuracy is 60%. Of the 795 people who would be predicted to sexually reoffend, 234 (29%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 71% of the time.

Table B-5
Accuracy Levels for Static-99 Cutoff of 8 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
Complete Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	38	52	90
Not Predicted to Reoffend	270	1,261	1,531
Totals	308	1,313	1,621

Sample BR = .19, TPR (Sensitivity) = .12, FPR = .04, Specificity = .96
Overall Accuracy at this BR = .8014, PPV at this BR = .42, NPV at this BR = .82

As shown in Table B-5 (above), with a base rate of .19 and a cut score of 8, overall accuracy is 80%. Of the 90 people who would be predicted to sexually reoffend, 38 (42%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 58% of the time.

Table B-6
Accuracy Levels for Static-99 Cutoff of 9 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
Complete Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	16	14	30
Not Predicted to Reoffend	292	1,299	1,591
Totals	308	1,313	1,621

Sample BR = .19, TPR (Sensitivity) = .05, FPR = .01, Specificity = .99
Overall Accuracy at this BR = .8112, PPV at this BR = .53, NPV at this BR = .82

As shown in Table B-6 (above), with a base rate of .19 and a cut score of 9, overall accuracy is 81%. Of the 30 people who would be predicted to sexually reoffend, 16 (52%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 48% of the time.

Table B-7
Accuracy Levels for Static-99 Cutoff of 13 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
Complete Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	308	1,313	1,621
Totals	308	1,313	1,621

Sample BR = .19, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .8100, PPV is undefined, NPV at this BR = .81

As shown in Table B-7 (above), with a base rate of .19 and a cut score of 13, overall accuracy is 81%. No one would be predicted to sexually reoffend. With detected sexual recidivism as our criterion standard, 81% of those predictions would be accurate.

In sum, for the 1,621 people included in the Complete Sample, 19% of them were detected to sexually recidivate within 10 years. Using detected sexual recidivism as our criterion standard, overall accuracy with the Static-99 is maximized with a cut score of 9. (This is slightly more accurate than predicting that no one will sexually re-offend.) With a cut score of 9, 30 (5%) of the people in the sample would be classified as predicted to sexually reoffend, and 16 (52%) of those predictions would be accurate.

C. Routine CSC Sample (October 2008), 10-Year Fixed Follow-up

Table B-8
Accuracy Levels for Static-99 Cutoff of 6 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
Routine CSC Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	16	46	62
Not Predicted to Reoffend	13	267	280
Totals	29	313	342

Sample BR = .085, TPR (Sensitivity) = .55, FPR = .15, Specificity = .85
Overall Accuracy at this BR = .8274, PPV at this BR = .26, NPV at this BR = .95

As shown in Table B-8 (above), with a base rate of .085 and a cut score of 6, overall accuracy is 83%. Of the 62 people who would be predicted to sexually reoffend, 16

(26%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 74% of the time.

Table B-9
Accuracy Levels for Static-99 Cutoff of 7 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
Routine CSC Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	12	18	30
Not Predicted to Reoffend	17	295	312
Totals	29	313	342

Sample BR = .085, TPR (Sensitivity) = .41, FPR = .06, Specificity = .94
Overall Accuracy at this BR = .8977, PPV at this BR = .40, NPV at this BR = .95

As shown in Table B-9 (above), with a base rate of .085 and a cut score of 7, overall accuracy is 90%. Of the 30 people who would be predicted to sexually reoffend, 12 (40%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 60% of the time.

Table B-10
Accuracy Levels for Static-99 Cutoff of 8 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
Routine CSC Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	3	10	13
Not Predicted to Reoffend	26	303	329
Totals	29	313	342

Sample BR = .085, TPR (Sensitivity) = .10, FPR = .03, Specificity = .97
Overall Accuracy at this BR = .8947, PPV at this BR = .23, NPV at this BR = .92

As shown in Table B-10 (above), with a base rate of .085 and a cut score of 8, overall accuracy is 89%. Of the 13 people who would be predicted to sexually reoffend, 3 (23%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 77% of the time.

Table B-11
Accuracy Levels for Static-99 Cutoff of 13 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
Routine CSC Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	29	313	342
Totals	29	313	342

Sample BR = .085, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .9152, PPV is undefined, NPV at this BR = .92

As shown in Table B-11 (above), with a base rate of .085 and a cut score of 13, overall accuracy is 91.5%. No one would be predicted to sexually reoffend, and 91.5% of those predictions would be accurate.

In sum, for the 342 people included in the 2008 Static-99 Routine CSC Sample with 10-year fixed follow-up, 8.5% of them were detected to sexually recidivate within 10 years. Using detected sexual recidivism as our criterion standard, overall accuracy with the Static-99 is maximized with a cut score of 13. No one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 91.5% of those predictions are accurate.

The next most accurate classification comes with a cut score of 7. Of the people in the sample, 30 (9%) would be classified as predicted to sexually reoffend, and 12 (40%) of those predictions would be accurate. With a cut score of 7, overall accuracy (90% correct) would be lower than predicting that no one would be detected to sexually recidivate within 10 years (91.5% correct).

D. High-Risk Sample (October 2008), 10-Year Fixed Follow-up

Table B-12
Accuracy Levels for Static-99 Cutoff of 7 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
High-Risk Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	56	71	127
Not Predicted to Reoffend	163	445	608
Totals	219	516	735

Sample BR = .298, TPR (Sensitivity) = .26, FPR = .14, Specificity = .86
Overall Accuracy at this BR = .6816, PPV at this BR = .44, NPV at this BR = .73

As shown in Table B-12 (above), with a base rate of .298 and a cut score of 7, overall accuracy is 68%. Of the 127 people who would be predicted to sexually reoffend, 56 (44%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 56% of the time.

Table B-13
Accuracy Levels for Static-99 Cutoff of 8 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
High-Risk Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	28	33	61
Not Predicted to Reoffend	191	483	674
Totals	219	516	735

Sample BR = .298, TPR (Sensitivity) = .13, FPR = .06, Specificity = .94
Overall Accuracy at this BR = .6952, PPV at this BR = .46, NPV at this BR = .72

As shown in Table B-13 (above), with a base rate of .298 and a cut score of 8, overall accuracy is 70%. Of the 61 people who would be predicted to sexually reoffend, 28 (46%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 54% of the time.

Table B-14
Accuracy Levels for Static-99 Cutoff of 9 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
High-Risk Sample, 10-Year Follow-up

	Detected to have reoffended	Not detected to have reoffended	totals
Predicted to Reoffend	12	12	24
Not Predicted to Reoffend	207	504	711
Totals	219	516	735

Sample BR = .298, TPR (Sensitivity) = .05, FPR = .02, Specificity = .98
Overall Accuracy at this BR = .7020, PPV at this BR = .50, NPV at this BR = .71

As shown in Table B-14 (above), with a base rate of .298 and a cut score of 9, overall accuracy is 70%. Of the 24 people who would be predicted to sexually reoffend, 12 (50%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 50% of the time.

Table B-15
Accuracy Levels for Static-99 Cutoff of 10 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
High-Risk Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	3	6	9
Not Predicted to Reoffend	216	510	726
Totals	219	516	735

Sample BR = .298, TPR (Sensitivity) = .01, FPR = .01, Specificity = .99
Overall Accuracy at this BR = .6980, PPV at this BR = .33, NPV at this BR = .70

As shown in Table B-15 (above), with a base rate of .298 and a cut score of 10, overall accuracy is 70%. Of the 9 people who would be predicted to sexually reoffend, 3 (33%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 67% of the time.

Table B-16
Accuracy Levels for Static-99 Cutoff of 13 or Higher
Detailed Recidivism Tables Static-99 (October 2008)
High-Risk Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	219	516	735
Totals	219	516	735

Sample BR = .298, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .7020, PPV is undefined, NPV at this BR = .70

As shown in Table B-16 (above), with a base rate of .298 and a cut score of 13, overall accuracy is 70%. No one would be predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 70% of those predictions would be accurate.

In sum, for the 735 people included in the 2008 Static-99 High-Risk Sample with 10-year fixed follow-up, 29.8% of them were detected to sexually recidivate within 10 years. Overall accuracy with the Static-99 is maximized with a cut score of either 9 or 13. With a cut score of 9, 24 (3%) of the people in the sample would be classified as predicted to sexually reoffend, and 12 (50%) of those predictions would be correct. With a cut score of 13, no one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard and using a cut score of either 9 or 13, 70% of predictions are accurate.

II. Static-2002

Table B-17
Accuracy Levels for Static-2002 Cutoff of 8 or Higher, 5-Year Follow-up
Hanson et al. (2010)

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	123	289	412
Not Predicted to Reoffend	159	1,352	1,511
Totals	282	1,641	1,923

Sample BR = .147, TPR (Sensitivity) = .44, FPR = .18, Specificity = .82
Overall Accuracy at this BR = .7670, PPV at this BR = .30, NPV at this BR = .89

As shown in Table B-17 (above), with a base rate of .147 and a cut score of 8, overall accuracy is 77%. Of the 412 people who would be predicted to sexually reoffend, 123 (30%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 70% of the time.

Table B-18
Accuracy Levels for Static-2002 Cutoff of 10 or Higher, 5-Year Follow-up
Hanson et al. (2010)

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	41	72	113
Not Predicted to Reoffend	241	1,569	1,810
Totals	282	1,641	1,923

Sample BR = .147, TPR (Sensitivity) = .15, FPR = .04, Specificity = .96
Overall Accuracy at this BR = .8372, PPV at this BR = .36, NPV at this BR = .87

As shown in Table B-18 (above), with a base rate of .147 and a cut score of 10, overall accuracy is 84%. Of the 113 people who would be predicted to sexually reoffend, 41 (36%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 64% of the time.

Table B-19
Accuracy Levels for Static-2002 Cutoff of 12 or Higher, 5-Year Follow-up
Hanson, et al. (2010)

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	6	6	12
Not Predicted to Reoffend	276	1,635	1,911
Totals	282	1,641	1,923

Sample BR = .147, TPR (Sensitivity) = .02, FPR = .004, Specificity = .996
Overall Accuracy at this BR = .8534, PPV at this BR = .50, NPV at this BR = .86

As shown in Table B-19 (above), with a base rate of .147 and a cut score of 12, overall accuracy is 85%. Of the 12 people who would be predicted to sexually reoffend, 6 (50%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 50% of the time.

Table B-20
Accuracy Levels for Static-2002 Cutoff of 15 or Higher, 5-Year Follow-up
Hanson et al. (2010)

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	282	1,641	1,923
Totals	282	1,641	1,923

Sample BR = .147, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .8534, PPV is undefined, NPV at this BR = .85

As shown in Table B-20 (above), with a base rate of .147 and a cut score of 15, overall accuracy is 85%. No one would be predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 85% of those predictions would be accurate.

In sum, for the 1,923 people included in the Hanson et al. (2010) sample with 5-year fixed follow-up, 14.7% of them were detected to sexually recidivate within 10 years. Overall accuracy with the Static-2002 is maximized with a cut score of either 12 or 15. With a cut score of 12, 12 (<1%) of the people in the sample would be classified as predicted to sexually reoffend, and 6 (50%) of those predictions would be correct. With a cut score of 15, no one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 85% of predictions are accurate.

III. Static-2002R

A. Five-Year Sexual Recidivism Rates for Static-2002R: Routine Sample

Table B-21
Accuracy Levels for Static-2002R, Cutoff of 7 or Higher, 5-Year Follow-up
Routine Sample

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	12	64	76
Not Predicted to Reoffend	16	434	450
Totals	28	498	526

Sample BR = .053, TPR (Sensitivity) = .43, FPR = .13, Specificity = .87
Overall Accuracy at this BR = .8479, PPV at this BR = .16, NPV at this BR = .96

As shown in Table B-21 (above), with a base rate of .053 and a cut score of 7, overall accuracy is 85%. Of the 76 people who would be predicted to sexually reoffend, 12 (16%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 84% of the time.

Table B-22
Accuracy Levels for Static-2002R, Cutoff of 9 or Higher, 5-Year Follow-up
Routine Sample

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	4	12	16
Not Predicted to Reoffend	24	486	510
Totals	28	498	526

Sample BR = .053, TPR (Sensitivity) = .14, FPR = .02, Specificity = .98
Overall Accuracy at this BR = .9316, PPV at this BR = .25, NPV at this BR = .95

As shown in Table B-22 (above), with a base rate of .053 and a cut score of 9, overall accuracy is 93%. Of the 16 people who would be predicted to sexually reoffend, 4 (25%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 75% of the time.

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

As shown in Table B-23 (above), with a base rate of .053 and a cut score of 11, overall accuracy is 95%. Of the 2 people who would be predicted to sexually reoffend, 1 (50%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 50% of the time.

Table B-24
Accuracy Levels for Static-2002R, Cutoff of 14 or Higher, 5-Year Follow-up
Routine Sample

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	28	498	526
Totals	28	498	526

Sample BR = .053, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .9468, PPV is undefined, NPV at this BR = .95

As shown in Table B-24 (above), with a base rate of .053 and a cut score of 14, overall accuracy is 95%. No one would be predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 95% of those predictions would be accurate.

In sum, for the 526 people included in the Routine Sample with 5-year fixed follow-up, 5.3% of them were detected to sexually recidivate within 5 years. Overall accuracy with the Static-2002 is maximized with a cut score of either 11 or 14. With a cut score of 11, 2 (<1%) of the people in the sample would be classified as predicted to sexually reoffend, and 1 (50%) of those predictions would be correct. With a cut score of 14, no on would be predicted to sexually reoffend, and 95% of those predictions would be accurate.

B. Ten-Year Sexual Recidivism Rates for Static-2002R: Non-routine Sample

Table B-25
Accuracy Levels for Static-2002R, Cutoff of 8 or Higher, 10-Year Follow-up
Non-routine Sample

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	74	95	169
Not Predicted to Reoffend	135	462	597
Totals	209	557	766

Sample BR = .273, TPR (Sensitivity) = .35, FPR = .17, Specificity = .83
Overall Accuracy at this BR = .6998, PPV at this BR = .44, NPV at this BR = .77

As shown in Table B-25 (above), with a base rate of .273 and a cut score of 8, overall accuracy is 66%. Of the 169 people who would be predicted to sexually reoffend, 74 (44%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 56% of the time.

Table B-26
Accuracy Levels for Static-2002R, Cutoff of 10 or Higher, 10-Year Follow-up
Non-routine Sample

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	26	23	49
Not Predicted to Reoffend	183	534	717
Totals	209	557	766

Sample BR = .273, TPR (Sensitivity) = .12, FPR = .04, Specificity = .96
Overall Accuracy at this BR = .7311, PPV at this BR = .53, NPV at this BR = .74

As shown in Table B-26 (above), with a base rate of .273 and a cut score of 10, overall accuracy is 73%. Of the 49 people who would be predicted to sexually reoffend, 26 (53%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 47% of the time.

Table B-27
Accuracy Levels for Static-2002R, Cutoff of 12 or Higher, 10-Year Follow-up
Non-routine Sample

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	3	3	6
Not Predicted to Reoffend	206	554	760
Totals	209	557	766

Sample BR = .273, TPR (Sensitivity) = .01, FPR = .01, Specificity = .99
Overall Accuracy at this BR = .7272, PPV at this BR = .50, NPV at this BR = .73

As shown in Table B-27 (above), with a base rate of .273 and a cut score of 12, overall accuracy is 73%. Of the 6 people who would be predicted to sexually reoffend, 3 (50%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 50% of the time.

Table B-28
Accuracy Levels for Static-2002R, Cutoff of 14 or Higher, 10-Year Follow-up
Non-routine Sample

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	209	557	766
Totals	209	557	766

Sample BR = .273, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .7272, PPV is undefined, NPV at this BR = .73

As shown in Table B-28 (above), with a base rate of .273 and a cut score of 14, overall accuracy is 73%. No one would be predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 73% of those predictions would be accurate.

In sum, for the 766 people included in the Static-2002R Non-routine Sample with 10-year fixed follow-up, 209 (27.3%) of them were detected to sexually recidivate within 10 years. Overall accuracy is maximized with a cut score of 10. With a cut score of 10, 49 (6%) of the people in the sample would be classified as predicted to sexually reoffend, and 26 (53%) of those predictions would be correct.

The next most accurate classification (slightly less accurate than a cut score of 10) comes with a cut score of 12 or 14. With a cut score of 12, 6 (1%) of the people in the sample would be classified as predicted to sexually reoffend, and 3 (50%) of those pre-

dictions would be correct. With a cut score of 14, no one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 73% of those predictions are accurate.

C. Ten-Year Sexual Recidivism Rates for Static-2002R: High-Risk/Need Sample

Table B-29
Accuracy Levels for Static-2002R, Cutoff of 8 or Higher, 10-Year Follow-up
High-Risk/Need Sample

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	66	87	153
Not Predicted to Reoffend	123	366	489
Totals	189	453	642

Sample BR = .294, TPR (Sensitivity) = .35, FPR = .19, Specificity = .81
Overall Accuracy at this BR = .6729, PPV at this BR = .43, NPV at this BR = .75

As shown in Table B-29 (above), with a base rate of .294 and a cut score of 8, overall accuracy is 67%. Of the 153 people who would be predicted to sexually reoffend, 66 (35%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 65% of the time.

Table B-30
Accuracy Levels for Static-2002R, Cutoff of 10 or Higher, 10-Year Follow-up
High-Risk/Need Sample

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	26	21	47
Not Predicted to Reoffend	163	432	595
Totals	189	453	642

Sample BR = .294, TPR (Sensitivity) = .14, FPR = .05, Specificity = .95
Overall Accuracy at this BR = .7134, PPV at this BR = .55, NPV at this BR = .73

As shown in Table B-30 (above), with a base rate of .294 and a cut score of 10, overall accuracy is 71%. Of the 47 people who would be predicted to sexually reoffend, 26 (55%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 45% of the time.

Table B-31
Accuracy Levels for Static-2002R, Cutoff of 12 or Higher, 10-Year Follow-up
High-Risk/Need Sample

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	3	3	6
Not Predicted to Reoffend	186	450	636
Totals	189	453	642

Sample BR = .294, TPR (Sensitivity) = .02, FPR = .01, Specificity = .99
Overall Accuracy at this BR = .7056, PPV at this BR = .50, NPV at this BR = .71

As shown in Table B-31 (above), with a base rate of .294 and a cut score of 12, overall accuracy is 71%. Of the 6 people who would be predicted to sexually reoffend, 3 (50%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 50% of the time.

Table B-32
Accuracy Levels for Static-2002R, Cutoff of 14 or Higher, 10-Year Follow-up
High-Risk/Need Sample

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	189	453	642
Totals	189	453	642

Sample BR = .294, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .7056, PPV is undefined, NPV at this BR = .71

As shown in Table B-32 (above), with a base rate of .294 and a cut score of 14, overall accuracy is 71%. No one would be predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 71% of those predictions would be accurate.

In sum, for the 642 people included in the Static-2002R High-Risk/Need Sample with 10-year fixed follow-up, 189 (29.4%) of them were detected to sexually recidivate within 10 years. Overall accuracy is maximized with a cut score of 10. With a cut score of 10, 47 (7%) of the people in the sample would be classified as predicted to sexually reoffend, and 26 (55%) of those predictions would be correct.

The next most accurate classification (slightly less accurate than a cut score of 10) comes with a cut score of 12 or 14. With a cut score of 12, 6 (1%) of the people in the sample would be classified as predicted to sexually reoffend, and 3 (50%) of those pre-

dictions would be correct. With a cut score of 14, no one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 71% of those predictions are accurate.

IV. Static-99R

A. Routine Sample

Table B-33
Accuracy Levels for Static-99R Cutoff of 6 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
Routine Sample, 5-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	41	193	234
Not Predicted to Reoffend	104	2,068	2,172
Totals	145	2,261	2,406

Sample BR = .06, TPR (Sensitivity) = .28, FPR = .09, Specificity = .91
Overall Accuracy at this BR = .8766, PPV at this BR = .18, NPV at this BR = .95

As shown in Table B-33 (above), with a base rate of .06 and a cut score of 6, overall accuracy is 88%. Of the 234 people who would be predicted to sexually reoffend, 41 (18%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 82% of the time.

Table B-34
Accuracy Levels for Static-99R Cutoff of 9 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
Routine Sample, 5-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	7	14	21
Not Predicted to Reoffend	138	2,247	2,385
Totals	145	2,261	2,406

Sample BR = .06, TPR (Sensitivity) = .05, FPR = .01, Specificity = .99
Overall Accuracy at this BR = .9368, PPV at this BR = .33, NPV at this BR = .94

As shown in Table B-34 (above), with a base rate of .06 and a cut score of 9, overall accuracy is 94%. Of the 21 people who would be predicted to sexually reoffend, 7

(33%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 67% of the time.

Table B-35
Accuracy Levels for Static-99R Cutoff of 10 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
Routine Sample, 5-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	2	6	8
Not Predicted to Reoffend	143	2,255	2,398
Totals	145	2,261	2,406

Sample BR = .06, TPR (Sensitivity) = .01, FPR = .003, Specificity = .997
Overall Accuracy at this BR = .9381, PPV at this BR = .25, NPV at this BR = .94

As shown in Table B-35 (above), with a base rate of .06 and a cut score of 10, overall accuracy is 94%. Of the 8 people who would be predicted to sexually reoffend, 2 (25%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 75% of the time.

Table B-36
Accuracy Levels for Static-99R Cutoff of 13 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
Routine Sample, 5-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	145	2,261	2,406
Totals	145	2,261	2,406

Sample BR = .06, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .9397, PPV is undefined, NPV at this BR = .94

As shown in Table B-36 (above), with a base rate of .06 and a cut score of 10, overall accuracy is 94%. No one would be predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 94% of those predictions would be accurate.

In sum, for the 2,261 people included in the 2009 Static-99R Routine Sample with 5-year fixed follow-up, 145 (6%) of them were detected to sexually recidivate within 5 years. Overall accuracy with the Static-99R is maximized with a cut score of 13. No

one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 94% of those predictions are accurate.

The next most accurate classification (slightly less accurate than a cut score of 13) comes with a cut score of 10. With a cut score of 10, 8 (<1%) of the people in the sample would be classified as predicted to sexually reoffend, and 2 (25%) of those predictions would be correct.

B. Preselected-for-Treatment Sample

Table B-37
Accuracy Levels for Static-99R Cutoff of 6 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
Preselected-for-Treatment Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	27	51	78
Not Predicted to Reoffend	91	697	788
Totals	118	748	866

Sample BR = .136, TPR (Sensitivity) = .23, FPR = .07, Specificity = .93
Overall Accuracy at this BR = .8360, PPV at this BR = .35, NPV at this BR = .88

As shown in Table B-37 (above), with a base rate of .136 and a cut score of 6, overall accuracy is 84%. Of the 78 people who would be predicted to sexually reoffend, 27 (35%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 65% of the time.

Table B-38
Accuracy Levels for Static-99R Cutoff of 8 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
Preselected-for-Treatment Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	8	16	24
Not Predicted to Reoffend	110	732	842
Totals	118	748	866

Sample BR = .136, TPR (Sensitivity) = .07, FPR = .02, Specificity = .98
Overall Accuracy at this BR = .8545, PPV at this BR = .33, NPV at this BR = .87

As shown in Table B-38 (above), with a base rate of .136 and a cut score of 8, overall accuracy is 85%. Of the 24 people who would be predicted to sexually reoffend, 8 (33%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 67% of the time.

Table B-39
Accuracy Levels for Static-99R Cutoff of 9 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
Preselected-for-Treatment Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	1	5	6
Not Predicted to Reoffend	117	743	860
Totals	118	748	866

Sample BR = .136, TPR (Sensitivity) = .01, FPR = .01, Specificity = .99
Overall Accuracy at this BR = .8591, PPV at this BR = .17, NPV at this BR = .86

As shown in Table B-39 (above), with a base rate of .136 and a cut score of 9, overall accuracy is 86%. Of the 6 people who would be predicted to sexually reoffend, 1 (17%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 83% of the time.

Table B-40
Accuracy Levels for Static-99R Cutoff of 13 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
Preselected-for-Treatment Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	118	748	866
Totals	118	748	866

Sample BR = .136, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .8637, PPV is undefined, NPV at this base rate = .86

As shown in Table B-40 (above), with a base rate of .136 and a cut score of 13, overall accuracy is 86%. No one would be predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 86% of those predictions would be accurate.

In sum, for the 866 people included in the 2009 Static-99R Preselected-for-Treatment Sample with 10-year fixed follow-up, 118 (13.6%) of them were detected to sexually

recidivate within 10 years. Overall accuracy with the Static-99R is maximized with a cut score of 13. No one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 86% of those predictions are accurate.

The next most accurate classification (slightly less accurate than a cut score of 13) comes with a cut score of 9. With a cut score of 9, 6 (1%) of the people in the sample would be classified as predicted to sexually reoffend, and 2 (33%) of those predictions would be correct.

C. High-Risk/Need Sample

Table B-41
Accuracy Levels for Static-99R Cutoff of 6 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
High-Risk/Need Sample, 10-Year Follow-up
Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	85	118	203
Not Predicted to Reoffend	119	381	500
Totals	204	499	703

Sample BR = .29, TPR (Sensitivity) = .42, FPR = .24, Specificity = .76
Overall Accuracy at this BR = .6629, PPV at this BR = .42, NPV at this BR = .76

As shown in Table B-41 (above), with a base rate of .29 and a cut score of 6, overall accuracy is 66%. Of the 203 people who would be predicted to sexually reoffend, 85 (42%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 58% of the time.

Table B-42
Accuracy Levels for Static-99R Cutoff of 9 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
High-Risk/Need Sample, 10-Year

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	12	15	27
Not Predicted to Reoffend	192	484	676
Totals	204	499	703

Sample BR = .29, TPR (Sensitivity) = .06, FPR = .03, Specificity = .97
Overall Accuracy at this BR = .7055, PPV at this BR = .44, NPV at this BR = .72

As shown in Table B-42 (above), with a base rate of .29 and a cut score of 9, overall accuracy is 71%. Of the 27 people who would be predicted to sexually reoffend, 12 (44%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 56% of the time.

Table B-43
Accuracy Levels for Static-99R Cutoff of 10 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
High-Risk/Need Sample, 10-Year

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	4	6	10
Not Predicted to Reoffend	200	493	693
Totals	204	499	703

Sample BR = .29, TPR (Sensitivity) = .02, FPR = .01, Specificity = .99
Overall Accuracy at this BR = .7070, PPV at this BR = .40, NPV at this BR = .71

As shown in Table B-43 (above), with a base rate of .29 and a cut score of 10, overall accuracy is 71%. Of the 10 people who would be predicted to sexually reoffend, 4 (40%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 60% of the time.

Table B-44
Accuracy Levels for Static-99R Cutoff of 13 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
High-Risk/Need Sample, 10-Year

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	204	499	703
Totals	204	499	703

Sample BR = .29, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .7098, PPV is undefined, NPV at this BR is .71

As shown in Table B-44 (above), with a base rate of .29 and a cut score of 13, overall accuracy is 71%. No one would be predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 71% of those predictions would be accurate.

In sum, for the 703 people included in the 2009 Static-99R High-Risk/Need Sample with 10-year fixed follow-up, 204 (29%) of them were detected to sexually recidivate within 10 years. Overall accuracy with the Static-99R is maximized with a cut score of 13. No one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 71% of those predictions are accurate.

The next most accurate classification (slightly less accurate than a cut score of 13) comes with a cut score of 10. With a cut score of 10, 10 (1%) of the people in the sample would be classified as predicted to sexually reoffend, and 4 (40%) of those predictions would be correct.

D. Non-Routine Sample

Table B-45
Accuracy Levels for Static-99R Cutoff of 6 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
Non-Routine Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	118	185	303
Not Predicted to Reoffend	214	1,109	1,323
Totals	332	1,294	1,626

Sample BR = .204, TPR (Sensitivity) = .36, FPR = .14, Specificity = .86
Overall Accuracy at this BR = .7546, PPV at this BR = .39, NPV at this BR = .84

As shown in Table B-45 (above), with a base rate of .204 and a cut score of 6, overall accuracy is 75%. Of the 303 people who would be predicted to sexually reoffend, 118 (39%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 61% of the time.

Table B-46
Accuracy Levels for Static-99R Cutoff of 9 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
Non-Routine Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	13	21	34
Not Predicted to Reoffend	319	1,273	1,592
Totals	332	1,294	1,626

Sample BR = .204, TPR (Sensitivity) = .04, FPR = .02, Specificity = .98
Overall Accuracy at this BR = .7909, PPV at this BR = .38, NPV at this BR = .80

As shown in Table B-46 (above), with a base rate of .204 and a cut score of 6, overall accuracy is 79%. Of the 34 people who would be predicted to sexually reoffend, 13 (38%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 62% of the time.

Table B-47
Accuracy Levels for Static-99R Cutoff of 10 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
Non-Routine Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	4	8	12
Not Predicted to Reoffend	328	1,286	1,614
Totals	332	1,294	1,626

Sample BR = .204, TPR (Sensitivity) = .01, FPR = .01, Specificity = .99
Overall Accuracy at this BR = .7934, PPV at this BR = .33, NPV at this BR = .80

As shown in Table B-47 (above), with a base rate of .204 and a cut score of 10, overall accuracy is 79%. Of the 12 people who would be predicted to sexually reoffend, 4 (33%) of them were detected to have sexually reoffended. Predictions that a person in this sample would sexually recidivate would be wrong 67% of the time.

Table B-48
Accuracy Levels for Static-99R Cutoff of 13 or Higher
Detailed Recidivism Tables Static-99R (October 2009)
Non-Routine Sample, 10-Year Follow-up

	Detected to Have Reoffended	Not Detected to Have Reoffended	Totals
Predicted to Reoffend	0	0	0
Not Predicted to Reoffend	332	1,294	1,626
Totals	332	1,294	1,626

Sample BR = .204, TPR (Sensitivity) = .00, FPR = .00, Specificity = 1
Overall Accuracy at this BR = .7958, PPV is undefined, NPV at this BR = .80

As shown in Table B-48 (above), with a base rate of .204 and a cut score of 13, overall accuracy is 80%. No one would be predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 80% of those predictions would be accurate.

In sum, for the 1,626 people included in the 2009 Static-99R Non-Routine Sample with 10-year fixed follow-up, 332 (20.4%) of them were detected to sexually recidivate within 10 years. Overall accuracy with the Static-99R is maximized with a cut score of 13. No one is predicted to sexually reoffend. With detected sexual recidivism as the criterion standard, 80% of those predictions are accurate.

The next most accurate classification (slightly less accurate than a cut score of 13) comes with a cut score of 10. With a cut score of 10, 12 (1%) of the people in the sample would be classified as predicted to sexually reoffend, and 4 (33%) of those predictions would be correct.

received April 20, 2010; revision submitted July 3, 2010; accepted July 16, 2010

References

Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric assessment of dynamic risk factors for child molesters. *Sexual Abuse: A Journal of Research and Treatment, 19,* 347-367. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Bartosh, D. L., Garby, T., & Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy & Comparative Criminology, 47,* 422-438. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Bengtson, S. (2008). Is newer better? A cross-validation of the Static-2002 and the Risk Matrix 2000 in a Danish sample of sexual offenders. *Psychology, Crime & Law, 14,* 85-106. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Bigras, J. (2007). La prédiction de la récidive chez les délinquants sexuels [Prediction of recidivism among sex offenders]. *Dissertations Abstracts International, 68* (09). (UMI No. NR30941). Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Birnbaum, D., & Sheps, S. B. (1991). Validation of new tests. Infection Control and Hospital Epidemiology, 12(10), 622-624.

Boer, A. (2003). *Evaluating the Static-99 and Static-2002 risk scales using Canadian sexual offenders*. Unpublished master's thesis, University of Leicester, Leicester, United Kingdom. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Bonta, J., & Yessine, A. K. (2005). [Recidivism data for 124 released sexual offenders from the offenders identified in *The National Flagging System: Identifying and responding to high-risk, violent offenders* (User Report 2005-04). Ottawa: Public Safety and Emergency Preparedness Canada]. Unpublished raw data. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Brouillette-Alarie, S., & Proulx, J. (2008, October). Predictive and convergent validity of phallometric assessment in relation to sexual recidivism risk. Poster presented at the annual conference for the Association for the Treatment of Sexual Abusers, Atlanta, GA. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

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

Cortoni, F., & Nunes, K. L. (2007). *Assessing the effectiveness of the National Sexual Offender Program* (Research Report No. R-183). Unpublished report, Correctional Service of Canada. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

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.

Craissati, J., Bierer, K., & South, R. (2008). *What do sex offenders really get up to? Risk prediction, community failure and "sexually risky behaviours" in a nine year follow up study.* Unpublished manuscript. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

de Vogel, V., de Ruiter, C., van Beck, D., & Mead, G. (2004). Predictive validity of the SVR-20 and Static-99 in a Dutch sample of treated sex offenders. *Law & Human Behavior, 28,* 235-251.

Eher, R., Rettenberger, M., Schilling, F., & Pfäfflin, F. (2008). Failure of Static-99 and SORAG to predict relevant reoffense categories in relevant sexual offender subtypes: A prospective study. *Sexual Offender Treatment, 3*(1), 1-14.

Epperson, D. L. (2003). *Validation of the MnSOST-R, Static-99, and RRASOR with North Dakota prison and probation samples*. Unpublished Technical Assistance Report, North Dakota Division of Parole and Probation. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Frederick, R. I., & Bowden, S. C. (2009). The test validation summary. *Assessment, 16,* 215-236.

Haag, A. M. (2005). Recidivism data from 198 offenders detained until their warrant expiry date. From: Do psychological interventions impact on actuarial measures: An analysis of the predictive validity of the Static-99 and Static-2002 on a reconviction measure of sexual recidivism*. Dissertations Abstracts International, 66* (08), 4531B. (UMI No. NR05662)]. Unpublished raw data. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Hanson, R. K. (1998). What do we know about sex offender risk assessment? *Psychology, Public Policy, and Law, 4,* 50-72.

Hanson, R. K., Harris, A. J. R., Scott, T., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Corrections Research User Report No. 2007-05). Ottawa, ON, Canada: Public Safety Canada.

Hanson, R. H., Helmus, L., & Phenix, A. (2009). Static-99(R) and Static-2002(R): How to interpret and report in light of recent research. Downloaded April 8, 2010, from www.static99.org

Hanson, R. K., Helmus, L., & Thornton, D. (2010). Predicting recidivism among sexual offenders: A multi-site study of Static-2002. *Law and Human Behavior, 34(3),* 198-211.

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. H., Phenix, A., & Helmus, L. (2009, September). *Static-99(R) and Static-2002(R): How to Interpret and Report in Light of Recent Research*. Paper presented at a pre-conference workshop at the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX. Retrieved from www.static99.org

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

Hanson, R. K., & Thornton, D. (2003). *Notes on the development of the Static-2002* (User Report No. 2003-01). Ottawa, ON: Solicitor General Canada.

Harkins, L., & Beech, A.R. (2007). Examining the effectiveness of sexual offender treatment using risk band analysis. Unpublished manuscript. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

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.

Heilbrun, K., Douglas, K. S., & Yasuhara, K. (2009). Violence risk assessment: Core controversies. In J. K. Skeem, K. S. Douglas, & S. O. Lilienfeld (Eds.), *Psychological science in the courtroom: Consensus and controversy,* (pp. 333-357). New York: Guilford.

Helmus, L. M. D. (2007). *A multi-site comparison of the validity and utility of the Static-99 and Static-2002 for risk assessment with sexual offenders*. Unpublished B.A. thesis, Carleton University, Ottawa, Ontario, Canada.

Helmus, L. (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Helmus, L., Hanson, R. K., & Thornton, D. (2009). Reporting Static-99 in light of new research on recidivism norms. *The Forum, 21*(1), 38-45. Downloaded April 7, 2010, from the Norms page at www.static99.org

Heuer, L., Penrod, S., & Kattan, A. (2007). The role of societal benefits and fairness concerns among decision makers and decision recipients. *Law and Human Behavior, 31,* 573-610.

Hill, A., Habermann, N., Klusmann, D., Berner, W., & Briken, P. (2008). Criminal recidivism in sexual homicide perpetrators. *International Journal of Offender Therapy and Comparative Criminology, 52,* 5-20. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

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

Johansen, S. H. (2007). Accuracy of predictions of sexual offense recidivism: A comparison of actuarial and clinical methods. *Dissertations Abstracts International, 68* (03), B. (UMI No. 3255527). Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Knight, R. A., & Thornton, D. (2007). *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders* (Document No. 217618). Submitted to the U.S. Department of Justice. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Langström, N. (2004). Accuracy of actuarial procedures for assessment of sexual offender recidivism risk may vary across ethnicity. *Sexual Abuse: A Journal of Research and Treatment, 16*, 107-120. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Monahan, J., & Walker, L. (1994). *Social science and law: Cases and materials* (3rd ed.). Mineola, NY: The Foundation Press, Inc. Cited in L. Heuer, S. Penrod, & A. Kattan, (2007). The role of societal benefits and fairness concerns among decision makers and decision recipients. *Law and Human Behavior, 31,* 573-610.

Mossman, D. (2008). Analyzing the performance of risk assessment instruments: A response to Vrieze and Grove (2007). *Law and Human Behavior, 32,* 279-291.

Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussey, C. (2009). Rater (dis)agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law, 15,* 19-53.

Nicholaichuk, T. (2001, November). *The comparison of two standardized risk assessment instruments in a sample of Canadian Aboriginal sexual offenders.* Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Antonio, TX. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Phenix, A., Helmus, L., Hanson, R. K., & Thornton, D. (2009a). Reporting Static-2002R scores with 2009 recidivism norms (non-routine): A template for cases for which the norms for routine samples do not apply. Unpublished manuscript downloaded April 10, 2010, from Static-2002R Non-Routine Template (.doc / .pdf) as static2002rnon-routinetemplate.pdf at www.static99.org

Phenix, A., Helmus, L., Hanson, R. K., & Thornton, D. (2009b). Reporting Static-2002R scores with 2009 recidivism norms (routine samples): A template for cases for which the norms for routine samples apply. Unpublished manuscript downloaded July 3, 2010, from www.amyphenix.com/docs/Static2002R-routine-template-11-03-09.pdf)

Rettenberger, M., Matthes, A., Boer, D. P., & Eher, R. (2009). Prospective actuarial risk assessment: A comparison of five risk assessment instruments in different sexual offender subtypes. *International Journal of Offender Therapy and Comparative Criminology, 54*(2)*,* 169-186.

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.

Seto, M. C. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Assessment, 17*(2)*,* 156-167.

Swinburne Romine, R., Dwyer, S. M., Mathiowetz, C., & Thomas, M. (2008, October). *Thirty years of sex offender specific treatment: A follow-up Study*. Poster presented at the conference for the Association for the Treatment of Sexual Abusers, Atlanta, GA. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment

of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Ternowski, D. R. (2004). Sex offender treatment: An evaluation of the Stave Lake Correctional Centre Program. *Dissertations Abstracts International, 66* (06), 3428B. (UMI No. NR03201). Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Thornton, D., Helmus, L., & Hanson, R. K. (2009, October). *Does Static-2002 fully allow for the effects of age on release*? Paper presented at the 28th annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX.

Thornton, D., Hanson, L., and Helmus, L. (n.d.). Moving beyond the Standard Model for Actuarial Assessment for Sexual Offenders. Downloaded April 10, 2010, from www.ccoso.org/newsletter/**movingbeyondstandard**.doc

Vrieze, S. I., & Grove, W. M. (2008). Predicting sex offender recidivism. I. Correcting for item overselection and accuracy overestimation in scale development: II. Sampling error-induced attenuation of predictive validity over base rate information. *Law and Human Behavior, 32,* 266-278.

Wilson, R. J., Cortoni, F., & Vermani, M. (2007a). *Circles of support and accountability: A national replication of outcome findings* (Report No. R-185). Ottawa, ON: Correctional Service of Canada. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.

Wilson, R. J., Picheca, J. E., & Prinzo, M. (2007b). Evaluating the effectiveness of professionally-facilitated volunteerism in the community-based management of high-risk sexual offenders: Part two – A comparison of recidivism rates. *The Howard Journal, 46*, 327-337. Cited in L. Helmus, (2009). *Re-norming Static-99 recidivism estimates: Exploring base-rate variability across sex offender samples.* A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the degree Master of Arts. Department of Psychology, Carleton University.