Flying Blind with Naked Factors: Problems and Pitfalls in Adjusted-Actuarial Sex-Offender Risk Assessment[1]

Authors: Terence W. Campbell & Gregory DeClue

Terence W. Campbell, Ph.D., ABPP; Independent Practice; Sterling Heights, MI[2]
E-mail: tcampbell3920@comcast.net

Gregory DeClue, Ph.D., ABPP; Independent Practice; Sarasota, Florida

Abstract

Actuarial instruments are typically the centerpieces of evaluations pursuant to civil-commitment statutes for sex offenders. Almost as frequently as they rely on actuarial instruments, evaluators adjust actuarial data via weighing additional variables that are (presumably) correlated with recidivism. Typically, however, such variables are only weakly related to reoffending. This article reviews many problems and pitfalls undermining Adjusted Actuarial Assessment (AAA) and reports data demonstrating how ill-advised this procedure is. Publicly available data do not support a claim in a recent meta-analysis (Hanson and Morton-Bourgon, 2009, p. 7), "For all three measures, for all types of raters, and for all outcomes, the adjusted scores showed lower predictive accuracy than did the unadjusted actuarial scores." Based on available data, at its best, AAA neither increases nor decreases the accuracy of actuarial classification. At its worst, AAA dilutes actuarial accuracy.

Keywords: adjusted-actuarial assessment, base rates, risk variables, sex-offender recidivism

I. Introduction

In response to public outrage evoked by highly publicized sexual assaults, 19 states and the federal government have enacted sexually violent predator (SVP) statutes (Arizona, California, Florida, Illinois, Iowa, Kansas, Massachusetts, Minnesota, Missouri, New Hampshire, New Jersey, New York, North Dakota, Pennsylvania, South Carolina, Texas, Virginia, Washington, and Wisconsin). These laws mandate a civil hearing to determine whether a previously convicted sex offender should be civilly committed, and confined indefinitely, for treatment. Texas' SVP statute is unique in that it places committed offenders into supervised community settings (Fitch & Hammen, 2003).

Actuarial assessment has developed as the primary method for assessing the recidivism risk of sex offenders (Harris & Rice, 2003; Hanson, 2003). The most frequently used actuarial instruments include the Rapid Risk for Sex Offender Recidivism (RRASOR)

[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 extremely valuable criticism and guidance.

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

(Hanson & Thornton, 2000), the Static-99 (Hanson & Thornton, 2000), and the Minnesota Sex Offender Screening Tool-Revised (MnSOST-R) (Epperson, Kaul, & Hesselton, 1998). Actuarial instruments can address a combination of four (overlapping) variables: risk factors, protective factors, static factors, and dynamic factors.

Risk factors increase the risk of an offender sexually reoffending. Protective factors decrease the risk of reoffending. Static factors (e.g., number of previous convictions, number of previous charges, history of nonsexual violence, etc.) cannot change (or are relatively unchanging) over time. Dynamic factors (availability of family support, availability of employment, level of self-esteem, etc.) can change over time.

The available actuarial instruments address only a combination of static-risk factors. These instruments do not address the remaining three factor combinations. We know of no peer-reviewed research demonstrating improved accuracy (over that obtained from static-risk factors alone) by incorporating dynamic-risk, dynamic-protective, or static-protective factors. Most authors have not recommended the use of available actuarial instruments as "stand alone devices" (Hanson, 1998; Campbell, 2007; Doren, 2002; Sjöstedt & Långström, 2001; but see Harris and Rice, 2010; Quinsey, Harris, Rice, and Cormier, 1998, 2006; Rice, 2010; and Rice, Harris, and Hilton, 2010).

Rather than rely exclusively on actuarial estimates of recidivism risk, evaluators typically adjust actuarial data (Petrila & Otto, 2001). Adjusted Actuarial Assessment (AAA) involves a two-step procedure: (1) obtain an actuarial estimate of recidivism risk, (2) adjust that actuarial estimate—upward or downward—relying on risk factors that have been purported to have empirical support. The risk factors used for adjustments often utilize data reported by Hanson and Bussière (1998), or more recent data reported by Hanson and Morton-Bourgon (2005). Additionally, the www.static99.org website[3] appears to recommend the use of "Risk Factors outside of the Static-99": sexual deviance, treatment history, general criminality, intimacy deficits, sexual self regulation, general self regulation, attitudes tolerating sexual assault, cooperation with supervision, and diagnosed Cluster B Personality Disorder.

Reviewing AAA also necessitates addressing considerations of terminology. Campbell (2003) and Petrila and Otto (2001) have used the term "Adjusted Actuarial Assessment." Hanson (1998) previously used the term "Adjusted Actuarial Evaluation" (p. 66), but more recently referred to the "adjusted actuarial approach" (Hanson, Morton, & Harris, 2003, p. 159). Doren (2002) has used the term "clinically adjusted actuarial findings" (p. 162). Despite these terminological variations, there are no substantial differences between Adjusted Actuarial Assessment, Adjusted Actuarial Evaluation, Adjusted Actuarial Approach, and “clinically adjusted actuarial findings.” Despite the different terms used to designate it, they all refer to the same procedure.

[3] The website lists the following members of its Advisory Board: Amy Phenix, R. Karl Hanson, Andrew J. R. Harris, and David Thornton.

II. Shortcomings of Adjusted Actuarial Assessment

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

> Given the poor track record of clinical prediction, evaluators should, nevertheless, be exceedingly cautious about how much they adjust actuarial estimates. Many of the factors that clinicians intuitively believe to be related to sexual offense recidivism, such as denial and verbal statements of treatment motivation, have not been found to predict sexual offense recidivism over the long-term (p. 66).

In their 1998 study reporting correlational data between risk variables and sex offender recidivism, Hanson and Bussière advised:

> The predictive accuracy of most of the variables was also small (.10-.20), and no variable was sufficiently related [to recidivism] to justify its use in isolation. It was also unclear how best to combine the variables because their intercorrelations were unknown and would be expected to be rather high for certain variables (e.g., young and unmarried). Consequently, we do not recommend simply summing the items, using either unit weights or weights inferred from the [correlational] tables, to create risk scales. . . . Nevertheless, our results could be used to identify the factors worth considering in risk assessments (p. 358).

This article explains why the variables in the tables in their meta-analysis should not be used to adjust the results of an actuarial instrument. If the risk variables Hanson and Bussière identified cannot be used in isolation, nor combined, nor summed, nor weighted, it becomes logically impossible to rely on them for AAA. Further, the research reported in the 1998 meta-analysis contributed to the development of the Static-99. The risk factors reported in that meta-analysis were considered as the Static-99 was developed, so none of those research results should be considered to represent “risk factors outside the Static-99.” They are the very risk factors that were considered in the development of the Static-99. We address this further, as “double dipping,” in our Conclusions. Hanson (2002) subsequently wrote: "Much more research is required before adjustments to established actuarial measures using static factors can be done with any confidence" (p. 100).

Hanson and Morton-Bourgon (2005) commented on their data regarding risk factors and recidivism.

> Readers will note, however, that the predictive accuracy of most of the characteristics was small. Consequently, prudent evaluators need to consider a range of potential risk factors in an overall evaluation. The best methods for combining risk factors into an overall evaluation remain an active topic of scientific debate (p. 1159).

Not long ago, the validity of AAA was unknown. Petrila and Otto (2001) wrote: "Perhaps most important, there is no data on the validity of adjusted actuarial assessment of risk for sexual reoffending, the technique used by almost all professionals who employ actuarial tests in their assessments" (p. 3-8).

In 2002, when the validity of AAA had not been empirically researched, Doren contended that evaluators had little choice but to rely on it: "Without a sufficiently comprehensive mechanical procedure available, evaluators seem to be left with adding clinical judgments where the actuarial instruments leave off" (p. 161).

This is serious business. Almost all professionals who employ actuarial tests in their sexual-recidivism risk assessments (Petrila & Otto, 2001) have been using a procedure (AAA) that lacks empirical support.

A. Diluted Accuracy of AAA

Attempts at adjusting actuarial data too often dilute their accuracy. For example, Sawyer (1966) found that, when given actuarial predictions and asked to "improve" them via clinical adjustments, clinicians performed worse than the actuarial predictions alone. Harris, Rice, and Cormier (2002) used the Violence Risk Appraisal Guide (VRAG) to predict violent recidivism for 467 subjects acquitted as insane or unfit to stand trial. Over a five-year follow-up, the AUC value for the VRAG was .80. Commenting on their findings, Harris et al. reported: “Clinical judgments could not make incremental improvements upon violence predictions made by actuarial scores alone. This means that clinical judgment could not have been tapping a dimension of risk-related information not already assessed by the VRAG” (p. 391).

Though not addressing sex-offender recidivism per se, the previously cited studies (Harris et al., 2002; Sawyer, 1966) are important because of their methodological significance. Each study attempted to improve the accuracy of actuarial assessment via clinical judgment. In each of these studies, the attempts failed. This provides a context for the findings that AAA has consistently led to a decrease in accuracy of sex-offender recidivism risk assessments. Methodologically relevant data advise against AAA when assessing sex-offender recidivism risk. Accordingly, Quinsey, Harris, Rice, and Cormier (2006) insist that assessments of sex-offender recidivism risk must confine themselves to actuarial procedures, and actuarial procedures alone.

B. Inter-Rater Reliability

AAA also contends with the problem of how to best weight and combine the different factors (Hanson, 1998). Without an explicitly defined method for converting various risk factors into recidivism probabilities, different evaluators can reach very different conclusions when assessing the same offender. For example, Evaluator A attributes considerable significance to the offender's age, but Evaluator B views the offender's failure to complete treatment as more compelling. When they occur, these different conclusions

reflect low levels of inter-rater reliability between or among two or more evaluators assessing the same offender.

Many of the risk factors reported by Hanson and Bussière (1998) and Hanson and Morton-Bourgon (2005) correspond to ambiguous variables leading to pronounced differences of opinion between evaluators. Consider, for example, the differences between "standardized" and "global assessments" (Bierman, Nix, Maples, & Murphy, 2006). Standardized assessments are typically anchored at each pole by concrete examples of behavioral extremes. Global assessments, however, involve general impressions of some behavior that are more intuitive than not. Bierman et al. (2006) found that clinicians are more accurate when rating clients via standardized ratings compared to global impressions.

Assessing "anger problems," for example, involves difficult challenges of how to reliably define and identify variations in this factor. Absent a generally accepted definition of "anger problems," different evaluators will reach different conclusions when assessing this factor. As a result, the accuracy of identifying "anger problems" is obviously undermined. In turn, low levels of inter-rater reliability can only diminish the accuracy of AAA.

Murrie et al. (2009) found that an evaluator's adversarial allegiance (hired by the prosecution versus hired by the defense) could influence some assessment instrument scores in forensic evaluation. This effect might be more pronounced for risk factors that are vaguely defined (leaving more room for subjective interpretation).

C. AAA as a Paradox

The correlation of an actuarial instrument such as the Static-99 with sex offender recidivism is .33, which is a moderate correlation (Hanson & Thornton, 2000). When Hanson (1998) reanalyzed some of the data in Hanson and Bussière (1998), he found that the average correlation of AAA with sex offender recidivism was .23, which is a small correlation. The average correlation for unguided clinical judgment and sex offender recidivism was .07, a trivial correlation (Hanson, 1998). AAA therefore: (1) relies on methods that are weakly correlated with recidivism (clinical judgment), (2) to identify variables that bear a small correlation with recidivism (various risk factors), (3) in order to adjust actuarial estimates that are moderately correlated with recidivism. To say the least, this procedure creates a substantial margin for error.

Actuarial assessments were developed to avoid the many problems undermining clinical judgment when assessing recidivism risk. AAA therefore amounts to a curious paradox. In particular, AAA involves a "u-turn" away from actuarial assessment, and hastily retreating back into clinical judgment. Actuarial assessment minimizes reliance on clinical judgment. AAA, however, leads evaluators into reinserting clinical judgment. AAA can therefore be defined as an unstandardized, inconsistently applied, ad hoc procedure. In other words, evaluators relying on AAA respond in an improvised manner to the idiosyncratic features of a particular case. Without a well-defined decision-making

procedure for AAA, these improvised judgments can also be expected to vary inconsistently from one case to another.

Rather than reduce the margin of error associated with actuarial instruments, AAA can be expected to increase that margin via its excessive reliance on clinical judgment. To at least some degree, Doren (2002) recognizes the various problems undermining AAA. He advises: “The bottom line is that examiners need to remain cognizant of the very significant limitations to our current knowledge of appropriate clinical adjustments to actuarial results. Making clinical adjustments to actuarial results typically involves some untested assumptions. We at least should be aware of our assumptions when we make those adjustments” (p. 171).

Given the many limitations undermining AAA, awareness of its "untested assumptions" cannot overcome its limitations. Identifying these problems does not equate to solving them. "Cognizant" as evaluators may be when undertaking AAA, those efforts will not suffice. The data presented below (Tables 2-6) help to understand why evaluators fail more often than they succeed when they attempt to adjust actuarial data.

III. Data Integration and Diagnostic Signs

A. Data Integration

When resorting to AAA, evaluators often describe themselves as undertaking "data integration." Attempts at data integration assume it is preferable to avoid relying on any one source of information. Presumably, evaluators should instead consider the available data as a whole and identify relationships and patterns within that data. This kind of data integration, however, is exceedingly difficult. Evaluators must decide what information is most relevant, how to obtain that information, and how to prioritize and integrate the different information.

Given the many difficulties challenging attempts at data integration, Faust (1986) advised: “Among the approaches that clinicians can use in their attempts to maximize diagnostic accuracy (whether it involves assignment to a diagnostic category or description of the individual), that of data integrator is nearly the worst” (p. 423).

B. Diagnostic Signs

Using various risk factors for AAA amounts to relying on the presumed significance of diagnostic signs. In what now amounts to a citation classic, Dawes, Faust, and Meehl (1989) reviewed the problems of diagnostic signs and their significance. They wrote:

> For example, suppose that about half of the adolescents appraised for a history of juvenile delinquency show subtle electroencephalographic (EEG) abnormalities. Based on these concurrences, the clinician may come to consider EEG abnormality a sign of delinquency or may conclude that delinquency is associated with brain dysfunction (p. 1671).

Dawes et al. then detailed the pitfalls associated with relying on risk factors such as these.

> One cannot determine, however, whether a relation exists unless one also knows whether the sign occurs more frequently among those with, versus those without, the condition. For example, to determine whether EEG abnormality is associated with delinquency, one must also know the frequency with which delinquents do not obtain EEG abnormalities and the frequencies with which nondelinquents do and do not obtain EEG abnormalities (p. 1672).

Most importantly, Dawes et al. further explained:

> Further, even should a valid relation exist, one cannot determine the sign's actual utility unless one knows: (i) how much more frequently it occurs when the condition is present than when it is absent and (ii) the frequency of the condition ... If the condition is infrequent, then positive identifications based on the sign's presence can even be wrong in most cases, for most individuals who display the sign will not have the condition (p. 1672).

Table 1 summarizes all possible outcomes associated with risk factors and sexual reoffending. If the frequency of offenders in cell B - factor present, but does not reoffend - exceeds the frequency of offenders in cell A - factor present and does reoffend - relying on that factor for risk assessment purposes is ill-informed and therefore ill-advised.

Table 1. Summarizing Risk Factor Frequencies for Recidivists and Non-Recidivists

	Reoffends	Does Not Reoffend
Factor Present	A	B
Factor Absent	C	D

Carefully reviewing the frequencies of presumed risk factors, distributed across samples of recidivists and non-recidivists, leads to some interesting outcomes.

C. Frequencies of Risk Variables

Scalora and Garbin (2003) reported treatment-related variables associated with recidivism, and non-recidivism, for a sample of 194 convicted child molesters. For this sample, 48 (24.7%) reoffended, and 146 (75.3%) did not reoffend, over an average follow up of approximately 54 months. This cognitive-behavioral treatment program, housed in a secure, inpatient facility, addressed offender responsibility, victim empathy, human sexuality, everyday living skills, and relapse prevention. The offenders treated in this program averaged a 28.3 months' stay. A multidisciplinary treatment team assessed each offender's response to the treatment program.

Scalora and Garbin (2003) reported the following treatment factors and their relation to recidivism (Table 2). Consider, for example, the "Untreated" group of offenders. On a percentage basis, recidivism occurred more frequently for this group (72.9%) compared to the nonrecidivism group (55.5%). These data might lead one to conclude that "untreated" status for this sample can rule in recidivism risk. In real numbers, however, there were 81 untreated offenders who did not reoffend compared to 35 untreated offenders who did reoffend. Concluding that an "untreated" offender is at risk for recidivism results in a 69.8% rate of false positive classifications (81/116).

Table 2. Treatment Factors and Recidivism (Scalora and Garbin, 2003)

	Recidivist		Non-recidivist	
Variable	%	(n)	%	(n)
Treatment outcome**				
Untreated	72.9	(35)	55.5	(81)
Tx drop out/terminated	10.4	(5)	6.2	(9)
Complete unsuccessfully	14.6	(7)	16.4	(24)
Successfully treated	2.1	(1)	21.9	(32)
Partial/full responsibility*	53.2	(25)	69.7	(101)
Multiple paraphilia history*	20.8	(10)	33.0	(54)

*p<.05, **r<.01

Scalora and Garbin interpreted their data as demonstrating: "Recidivism is significantly related to quality of treatment involvement. . . . Successfully treated offenders were significantly less likely to subsequently reoffend" (p. 309).

Assume that an evaluator applies the above data to adjust the results of an actuarial risk assessment, the Static-99. Begin with the example of a person who was “successfully treated” with the same type of treatment as that delivered in the above study. Of the 33 people who successfully completed treatment, 32 (97%) were not detected to sexually re-offend. If similar results were obtained consistently for large samples of sex offenders, then why use an actuarial instrument at all? If such results were widely found for large groups of sex offenders, an evaluator could over-ride (ignore) the actuarial score for anyone who successfully completed this type of sex-offender treatment, and report that the likelihood of detected sexual re-offense for all those successfully treated sexual offenders is very low.

Now, with the same data set, imagine that an evaluator has decided to adjust an actuarial instrument, the Static-99, for someone who was not “successfully treated.” Collapsing the small groups reported above ("untreated,” “drop out/terminated,” and "complete unsuccessfully"), we find that 47 reoffended, and 114 did not reoffend. Most people in those three categories were not detected to sexually re-offend, so how much, if any, should an evaluator adjust an actuarial risk assessment when it is determined that

a person had not completed sex-offender treatment? Note that the overall detected sexual recidivism rate for the entire sample was 24.7%, and the detected sexual recidivism rate for the combined group of those not classified as "successfully treated" was 29.2% (47 of 161). If similar findings were reported across a wide range of "not successfully treated" sex offenders, what should an evaluator make of that 4.5% difference in recidivism rate?

It is tempting to imagine that, if there were widespread findings exactly like Scalora and Garbin reported, an evaluator could (a) ignore the actuarial instrument and report "low risk" for anyone who successfully completed treatment and (b) ignore the treatment results and use the actuarial instrument for everyone not classified as "successfully treated." However, if there really were widespread findings exactly like Scalora and Garbin reported, then the actuarial instruments themselves could be revised, with an item for treatment outcome (perhaps with two scores, 0 for not successfully treated and -10 for successfully treated!).

There is also the possibility that a risk factor based on treatment response would wash out of the statistical analysis as other variables are combined. Nunes and Corontoni (2008, p. 24) examined "the extent to which general criminality and sexual deviance are associated with dropout or expulsion from a sex-offender treatment program" for 100 subjects and found "The general criminality items of the Static-99 were significantly associated with dropout/expulsion."

For now, Scalora and Garbin's (2003) findings suggest that there could be a substantial impact of successful completion of sex-offender treatment, but offer essentially no support for increasing estimated risk for those who are not successfully treated. Nunes and Corontoni's (2008) findings suggest that an unknown amount of the predictive power of treatment response is already built into the Static-99's general criminality items, which would reduce or eliminate any potential gains from adjusting Static-99-based estimates of risk.

It would be hard to overstate the need for evaluators to consider sample size in deciding whether there are sufficient data to warrant an adjustment to an actuarial instrument based on research involving extraneous variables. Note that Scalora and Garbin's entire sample was 194 people. When an auto insurance company recently decided to change the rates they charge their customers due to new data regarding speeding rates of various demographic groups, the change was "based on a total sample size of 2.74 million" (Karouski, 2010). The auto-insurance sample size that showed a need to change rates is 14,421 times as large as Scalora and Garbin's sample.

D. Low Self-Esteem as a Presumed Risk Factor

Consider an example of a clinical factor that an evaluator might consider using to adjust an actuarial-based risk assessment. Thornton, Beech, and Marshall (2004) obtained self-esteem data on a pretreatment basis from 225 sex offenders. The follow-up period for this sample ranged between 4 and 6 years. The recidivism base rate for this sample

was 9.33%. Thornton et al. reported: “Lower levels of self-esteem were associated with higher sexual recidivism rates” (p. 587).

Thornton and his colleagues moreover claimed: "The results give clear support to the hypothesis that low self-esteem prior to treatment will be associated with higher sexual recidivism rates" (p. 593).

As evaluators come across a study like Thornton et al. (2004) showing statistically significant differences in a measurable variable (in this case, self-esteem) between groups of people who were detected to have sexually reoffended and those who were not so detected, it may be tempting to think that such research should guide risk assessment. If Thornton et al. found statistically significant differences, why not adjust the actuarial assessment of the next person you evaluate?

First, there is the sample size; not 2.74 million people as in the insurance study mentioned above, but 225 people. Statistical significance is important, but this is still a small group of people in one sample, who may be different from the next person to be evaluated in a different setting.

Next, consider what Thornton et al.’s data allow one to say about a person in their own study. For their combined samples, Thornton and colleagues reported the data found in Table 3.

Table 3. Levels of Self-esteem and Recidivism

	Recidivist	Non-recidivist
Self-esteem		
Low	12	59
Moderate	6	68
High	3	77

In percentage terms, 16.9% of offenders exhibiting low self-esteem reoffended (12/71). For the moderate self-esteem group, 8.1% reoffended (6/74); and for the high self-esteem group, 3.7% reoffended (3/80). The raw data reported in this study allow computing a 2 X 2 contingency table summarizing its results. For purposes of this 2 X 2 table, moderate self-esteem and high self-esteem were collapsed into one category. Using "Low self-esteem" to rule in recidivism, and using "Moderate-High self-esteem to rule-out recidivism leads to the following outcomes.

Table 4. Self-esteem, Recidivism, and Classification Accuracy

	Reoffends	Does not reoffend
Low self-esteem	12	59
Moderate-high self esteem	9	145

If relying on "low self-esteem" to rule in recidivism risk, the frequency of false positive classifications is 83% (59/71). That is, even in Thornton et al.'s own sample, if we know that a person had low self-esteem prior to entering treatment, should we consider that he is likely to be detected to have sexually reoffended at some point after treatment? No, we should not, because 83% of those with low self-esteem were not detected to have sexually reoffended. The 17% detected sexual-recidivism rate for people with low self-esteem is greater than the 9% base rate of sexual reoffending for this study, but this study does not show that self-esteem is such a powerful variable that an evaluator should use it to adjust the estimated risk from an actuarial instrument.[4]

If relying on "moderate-high self-esteem" to rule out recidivism risk, the frequency of true negative classifications would be 94% (145/154). If relying on the base rate alone for this sample, the overall classification accuracy of concluding that none of these offenders would reoffend is 90.6% (204/225). Consequently, using "moderate-high self-esteem" to rule out recidivism risk would allow a level of classification accuracy slightly greater than relying on the base rate, even for the subjects in this sample. Although some evaluators might wish to take a less extreme approach than *ruling out* sexual-recidivism risk on the basis of moderate-high self esteem, this analysis helps to illustrate that self-esteem is not a strong enough risk factor that it should be used to adjust the actuarial-instrument-based risk estimate of the next person an evaluator considers for possible civil commitment.

IV. HANSON AND MORTON-BOURGON (2005)

In their 2005 meta-analysis of risk factors, Hanson and Morton-Bourgon relied on 82 different studies including 29,450 sex offenders. Hanson and Morton-Bourgon reported data for six studies identifying "sexual deviance" in dichotomous terms (deviance present or deviance absent). For the 261 sexual recidivists, 65.5% exhibited sexual deviance compared to 42.3% of the 749 non-recidivists.0

If 65.5% of the 261 recidivists exhibited deviancy, this equates to 171 offenders. If 42.3% of the 749 non-recidivists exhibited deviancy, this equates to 317 offenders (see Table 5). This is a significant difference across multiple samples. But, if one were to use sexual deviancy alone to rule in recidivism risk, that would result in a false-positive percentage of 65% (317 of 488 offenders exhibiting deviancy do no reoffend[5]).

[4] Recall that some actuarial instruments, including the Static-99, have been tested with larger numbers of subjects over multiple samples, and only items that provide incremental validity are retained in the final instrument.

[5] Using detected sexual recidivism as the criterion variable.

Table 5. Frequencies of "Sexual Deviancy"
Found in Samples of Recidivists and Non-recidivists

	Reoffends	No Reoffending
Factor present	171	317
Factor absent	90	432

Table 5 demonstrates the necessity of carefully defining the question at hand when considering risk factors. If asking, "given that Mr. Smith is a recidivist, does he likely exhibit sexual deviancy?" the answer is yes. However, this is not the question SVP evaluators encounter. SVP assessments cannot, or at least should not, begin with the presumption that a given offender will reoffend. Instead, SVP evaluators are asked, "Given that Mr. Smith exhibits sexual deviance (or any other risk factor), will he likely reoffend?"[6]

Above all else, Table 5 demonstrates the importance of base rates. For example, if you know that I (the first author) reside in Michigan, you can accurately conclude that I live east of the Mississippi River. On the other hand, knowing only that I live east of the Mississippi River does not allow you to conclude that I reside in Michigan. Correspondingly, if you know that Mr. Smith is a sexual reoffender, you can usually conclude that he exhibits sexual deviance. However, if all you know is that Mr. Smith exhibits sexual deviance, you cannot conclude that he will sexually reoffend. Relying on risk factors to rule in recidivism is similar to relying on behavioral indicators to identify sexually abused children. In both instances, false positive classifications occur more frequently than true positives because of base rate considerations (Friedrich, 2005; Wood, 1996).

For the Hanson and Morton-Bourgon (2005) "sexual deviancy" sample, the base rate of sexual reoffending was 25.8%. More recently available data (Helmus, 2009; Helmus, Hanson, & Thornton, 2009) indicate that this is an atypically high base rate. For lower recidivism base rates, the frequencies of false positive classifications will be even greater if attempting to rule in recidivism risk vis-à-vis risk factors outside an actuarial instrument.

Unfortunately, the data reported in Hanson and Morton-Bourgon (2005) do not provide enough detail to allow the kind of 2 X 2 contingency summary found in Table 5 for other risk factors. In other words, the tables do not report the frequencies with which risk factors such as sexual preoccupations, antisocial personality disorder, self-regulation problems, employment instability, hostility, etc., occur in samples of recidivists, compared to non-recidivists. Knowing a correlation or effect size is insufficient. Evaluators wishing to use these factors for AAA should have at least enough data to compute positive predictive values (with confidence intervals). Even with enough data to compute positive predictive values, clinical adjustments without detailed knowledge of interactive

[6] In other words, the evaluator should be interested in the tool's positive predictive value. We address this further in another paper (currently under review).

effects among variables are guesses. Without enough knowledge to compute positive predictive values, clinical adjustments are hardly even educated guesses.

Another consideration in deciding whether to use a variable such as sexual deviancy to adjust an actuarial-instrument-based risk assessment is whether that variable has already been included in some way in the risk-assessment tool. There are items in the Static-99 that are designed to tap sexual deviancy, so it would be “double dipping” to adjust a Static-99-based risk assessment on the basis of whether or not the person was identified to show a sexual deviancy.

V. FALLIBLE HUMAN BEINGS

Nobel Prize winner Daniel Kahneman has discussed some of the natural ways that people think about probability and statistics: “People jump to statistical conclusions on the basis of very weak evidence. We form powerful intuitions about trends and about the replicability of results on the basis of information that is truly inadequate” (Jaffe, 2004, p. 25).

Kahneman continued to comment about "overconfident experts," indicating, "What you find is a great deal of confidence in the presence of very poor accuracy. . . . So the confidence people have is not a good indication of how accurate they are" (p. 25).

Kahneman further clarified: “When people make decisions they tend to suppress alternative interpretations. We become aware only of a single solution—this is a fundamental rule in perceptual processing. All the other solutions that might have been considered by the system—and sometimes we know that alternative solutions have been considered and rejected—we do not become aware of. So consciousness is at the level of a choice that has already been made” (p. 26).

Davis and Follette (2002) explained that relying on variables such as risk factors to predict some outcome assumes: "This person has all the characteristics of a murderer/bad parent/thief [or recidivist], therefore, (s)he must (will) be one" (p. 134).

Davis and Follette described the above assumptions as amounting to "intuitive profiling," explaining: “Essentially the use of intuitive profiling to assess guilt [or recidivism risk] relies on the following logic: 'If persons who commit embezzlement are likely to be in debt, then persons who are in debt are likely to be embezzlers,' or 'If most A's are B's, then most B's are A's.' Logically, of course, these conclusions are erroneous” (p. 134).

How would intuitive profiling affect AAA? It could lead evaluators into automatic thinking patterns such as the following: If sex offender recidivists (A's) are frequently untreated (B's), then most sex offenders who are untreated (B's) are recidivists (A's).

When an evaluator decides to open the door by using clinical judgment to adjust an actuarial assessment, any of the well-documented fallacies of human thinking might

rush in because, as Lilienfeld (in press, p. 2) puts it, “Scientific thinking does not come naturally to any of us.”

VI. SOME RECENT DEVELOPMENTS

A. “Risk Factors outside of the Static-99”

During a February 2009 deposition,[7] Dr. Amy Phenix made it clear and evident that she now rejects relying on risk factors outside the Static-99 (New Hampshire v. William Ploof). Phenix's opinion is significant because she serves on the Advisory Board of the www.static99.org website. During this deposition, the following Q and A exchange occurred:

> Dr. P: Karl Hanson and Andrew Harris wrote an article that described best practices with risk assessment that's published on their website [www.static99.org], and in one of the journals in my field, that absolutely recommended a clinically adjusted actuarial approach.
>
> Attorney: And now, just -- to be clear, now the recommendation is contrary to that, right?
>
> Dr. P: The recommendation is contrary to overriding a risk assessment, an actuarial instrument with risk factors outside the -- whatever instrument you're using (p. 77-78).

As her testimony continued, Phenix made it clear and evident that she considers any risk factors outside the Static-99 as obsolete:

> Attorney: I just came across this worksheet of risk factors to assess outside the Static-99. Can you see that?
>
> Dr. P: Yes, I know it.
>
> Attorney: It's on the static99.org web site. Is this obsolete now?
>
> Dr. P: Yes. I took it off of my web site. Apparently, they haven't taken it off of that one yet (p. 169).

This presents a dilemma for evaluators. In January 2009, one of the developers of the Static-99 testified under oath that the “Risk Factors outside of the Static-99” document is obsolete, but that “they” have not yet taken it off the www.static99.org website. At the time of this writing more than a year later (March 2010), that document is still on the website (and Phenix is still one-fourth of the Advisory Board for the website).

[7] http://web.me.com/gregdeclue/Site/Phenix_depo.html

Also on the website is a six-page document titled "Frequently Asked Questions," that ends with "The above rules apply equally to Static-99, Static-99R, Static-2002, and Static-2002R." Apparently, evaluators are currently (May 2010) encouraged to consider the 27 "rules" delineated in these Frequently Asked Questions as authoritative for all four of those instruments. The response to Question 12 specifically addresses adjusting these actuarial instruments:

> 12. What does a "departure" from the score mean? Can I adjust the score?
>
> Static-99 scores cannot be adjusted, or overridden, without the corresponding recidivism estimates being voided. The Static-99 is not intended to be a comprehensive risk assessment instrument, and as such, some factors related to risk are not included because they failed to predict recidivism in this particular model. However, that does not mean that those factors should not be considered in an overall assessment of risk.
>
> A "departure" refers to extenuating or acute risk factors that are not accounted for by the instrument that presently outweigh the level of risk arrived at via the actuarial assessment. For example, an acute risk would be someone who scores as low risk (0 or 1) but has stated an intention to reoffend. An empirically-based example would be an offender with a high risk score who has severe health issues and no access to potential victims. When the assessor feels that a departure is necessary, he or she should include the empirically-based or acute reasons in writing along with the assessment.

At least one part of this answer is clear. Evaluators are not encouraged to change the score of the risk-assessment instrument (Static-99, Static-99R, Static-2000, Static-2000R) based on outside factors.

A second part of this answer is also clear. The FAQ at www.static99.org recognizes that[8] factors outside the actuarial risk-assessment instrument may be so powerful that an evaluator should override (mention but disregard) the actuarial-instrument-based risk assessment and rely instead on the external factor.

It is not clear, though, what is recommended by www.static99.org for less extreme cases. Does www.static99.org currently recommend use of the "Risk Factors outside of the Static-99," which are still available for download[9] on the home page at www.static99.org, but which Phenix said were already obsolete over a year ago? The quote above recommends a departure from the risk derived from the actuarial instrument when there are "extenuating or acute risk factors that are not accounted for by the instrument that presently outweigh the level of risk arrived at via the actuarial assessment." However, as described in this article for some examples, the "Risk Factors out-

8 At least in some extreme cases (see the examples in the above quote).
9 The document is listed along with the coding rules, coding forms, and other essential documents.

side of the Static-99" are not so powerful that they should be considered to outweigh the level of risk arrived at via actuarial assessment.

B. Hanson and Morton-Bourgon (2009)

In recent years, sex-offender recidivism research and practice has been guided by influential meta-analyses. One recent meta-analysis is Hanson and Morton-Bourgon's (2009) "The Accuracy of Recidivism Risk for Sexual Offenders: A Meta-Analysis of 118 Prediction Studies. Hanson and Morton-Bourgon (2009, p. 7) wrote:

> Three studies examined the difference between actuarial scores and adjusted actuarial risk ratings (Gore, 2007; Hanson, 2007; Vrana, Sroga, & Guzzo, 2008). In these studies, evaluators were required to complete an actuarial risk tool and then were allowed to adjust the final risk rating on the basis of factors external to the actuarial tool. All three studies were prospective, and evaluators completed the ratings as part of their routine procedures. In two studies, the raters were probation officers (Hanson, 2007; Vrana et al., 2008), and in the other study, the raters were either psychologists or correctional staff (Gore, 2007). For all three measures, for all types of raters, and for all outcomes, the adjusted scores showed lower predictive accuracy than did the unadjusted actuarial scores.[10]

That conclusion (the last sentence in the above quote) is stated strongly and without equivocation. We were surprised to find that it is not supported by publicly available data.

Our reading of Gore (2007) is that adjusting scores failed to exceed the accuracy of actuarial scores, not that adjusting led to lower scores. Regarding clinical adjustments to the Minnesota Sex Offender Screening Tool (MnSOST-R), Gore (2007, p. 60) wrote, "The primary purpose of this study was to determine the impact of clinical overrides on predictive accuracy. The argument supporting an adjusted actuarial approach rests on its presumed greater accuracy relative to a purely actuarial approach. Because of this argument, justification for an adjusted actuarial approach would require that it result in significantly greater predictive accuracy, not just in an equivalent level of accuracy. Other factors supporting a requirement of significantly greater accuracy for an adjusted actuarial approach are its greater cost, in terms of time and personnel, and the fact that this approach at least opens the door for inappropriate overrides." She found (p. 61),

---

10 Hanson and Morton-Bourgon (2009, p. 7) considered studies comparing structured professional judgment (SPJ) versus mechanical/actuarial assessment separately: "We identified five studies in which the evaluators rated a predetermined set of items and then formed an overall evaluation of risk based on either (a) professional judgment or (b) summing the items. . . . The results of the procedures were similar. Three studies favored professional judgment, and two studies favored the simple sums. In most cases, the differences between the approaches was [sic] not large enough to be meaningful." One of those studies, de Vogel, de Ruiter, van Beck, and Mead, 2004, reported greater accuracy for structured professional judgment than actuarial assessment, but the judgment ratings had low inter-rater reliability (intraclass correlation coefficient [ICC] = .48). SPJ is not a primary focus of this paper; we return to it briefly in our conclusions.

"ROC[11] analyses failed to reveal any statistically significant differences in overall accuracy between the purely actuarial approach (MnSOST-R risk levels) and the adjusted actuarial approach (clinically overridden risk levels). In fact, when all overrides were considered, irrespective of direction, the accuracy of the adjusted actuarial approach of both psychologists and the ECRC [End of Confinement Review Committee] failed to even nominally exceed the accuracy of the actuarial approach. Nominally, the lowest ROC was for the ECRC's, which had the greatest number of overrides."

Slide 17 of Hanson (2007) briefly reports that the AUCs for "Sexual Recidivism," "Sexual Recidivism + Sex Breaches," and "Any Violent" were all lower for "Static-99 + Override" than for "Static-99 Alone." Data that purportedly support those conclusions are not publicly available at this time (R. Karl Hanson, personal communication, May 15, 2010). Other data (not involving unstructured clinical adjustments to actuarial data) from the same research project have been published in a peer-reviewed journal (Hanson, Harris, Scott, and Helmus, 2007; see below).

As Vrana, Sroga, & Guzzo (2008) addressed the predictive validity of the LSI-OR among a sample of adult male sexual assaulters, they wrote, "It was also of interest to examine the effectiveness of the clinical Override, which is often used on this population to adjust an offender's risk level based on the professional judgement of the assessor. . . . This was exercised in 42.4% of the cases from the entire sample, 23.2% from the custody sample, and 60.6% from the community sample. . . . [For cases in which a clinical Override was employed,] all offenders' risk levels were increased due to the Override, except for two institutionalized offenders whose risk levels were lowered by one" (pp. 14-15). They found that for *general* and *violent* recidivism, the initial risk assessments based on the LSI-R were "more accurate at predicting recidivism than the overridden final risk levels, which were based on subjective professional judgment" (p. 19). Notably, Vrana et al. (2008) did not report findings that clinically adjusted scores showed lower predictive accuracy than did the unadjusted actuarial scores for *sexual* recidivism: "From the entire sample, only 3.0% of the sexual assaulters reoffended sexually. A chi-square did not show any significant differences in offenders across location and sexual recidivism. Due to the low rate of sexual recidivism, it was not possible to use this outcome measure in most of the subsequent analyses" (p. 14).

Consider again the quote from Hanson and Morton-Bourgon (2009, p. 7): "Three studies examined the difference between actuarial scores and adjusted actuarial risk ratings (Gore, 2007; Hanson, 2007; Vrana, Sroga, & Guzzo, 2008). For all three measures, for all types of raters, and for all outcomes, the adjusted scores showed lower predictive accuracy than did the unadjusted actuarial scores."

We have been unable to find publicly available data to support Hanson and Morton-Bourgon's conclusion. To summarize:

1. Gore (2007, p. 61) reported that her analyses failed to reveal any statistically significant differences in overall accuracy between the purely actuarial approach

and the adjusted actuarial approach, not that the clinically adjusted actuarial scores showed lower predictive accuracy.

2. For Hanson (2007), data that purportedly support the claim (that clinically adjusted scores showed lower predictive accuracy than did the unadjusted actuarial scores) have not been made available for public scrutiny (R. Karl Hanson, personal communication, May 15, 2010).
3. Vrana et al. (2008) reported that, for *general* and *violent* recidivism, clinically adjusted scores showed lower predictive accuracy than did the unadjusted actuarial scores, but they did not report such data regarding *sexual* recidivism.

At the time of this writing (May 2010), we find no direct, relevant, publicly available empirical support for a claim that, for the type of risk assessment central to SVP evaluations (sexual recidivism), clinically adjusted scores show lower predictive accuracy than unadjusted actuarial scores.

C. The Dynamic Supervision Project

Hanson, Harris, Scott, and Helmus (2007, p. i) described the results of the Dynamic Supervision Project, which "involved every Canadian province and territory and the states of Alaska and Iowa in a robust test of risk assessment methodologies. . . . All of the probation and parole officers scoring risk of reoffense for these community-based sexual offenders were trained in sexual offender risk assessment by attending a two-day training that focused on scoring actual case examples. . . . The sexual recidivism rate for this widely disparate group of community-based sexual offenders was 7.6% after three years (n = 790). . . . STABLE-2007 and ACUTE-2007 assessments were found to add predictive power above and beyond that available to assessments of static risk alone. This study provides evidence that trained community supervision officers can reliably score valid and useful sex offender risk assessments. Results of this nature, even taking into account the need for replication and cross-validation, suggest significant policy and practice implications for the community supervision of sexual offenders."

Hanson, Harris, Scott, and Helmus's (2007, p. 25) discussion includes: "Nevertheless, it was possible to improve the accuracy of risk assessments by using a structured approach to combine static, stable and acute characteristics into an overall evaluation of current risk. . . . The current study provided strong support for the value of using structured, empirically-based approaches to risk assessment. When the officers were given the opportunity to provide unstructured adjustments to the risk assessments (override, unique acute factors), the adjustments either did nothing (unique acute factors), or degraded the predictive accuracy of the original scores (override)."

How much did the Stable-2007 add to the accuracy of predictions based on the Static-99? Their Appendix 10 (p. 44) shows that, for sexual recidivism, the AUC[12] for Static-99 was .74 (95% CI = .67 to .80); for Static-99 combined with Stable-2007, AUC = .76 (95% CI = .69 to .82).

[12] Area under the curve. See Fan, Upadhye, and Worster (2006).

Consider these points about the Dynamic Supervision Project:

1. To date, this is the only peer-reviewed, published study we found that addresses whether use of the Stable-2007 and/or Acute-2007 improves accuracy of prediction of any of the instruments in the Static-99 fleet.[13]
2. Overall, for sexual recidivism, adding Stable-2007 to Static-99 increased AUC from .74 to 76. For the 613 Canadian offenders, adding Stable-2007 to Static-99 increased AUC from .76 to .79. For the 179 US offenders (Alaska and Iowa), adding the Stable-2007 increased AUC from .56 to .60.[14]
3. Generally, one would expect some shrinkage on cross validation.
4. Various unstructured adjustments decreased the accuracy of prediction or left accuracy unchanged.
5. This procedure was done by the offenders' probation officers while the offenders were on probation—not by an SVP evaluator while the offenders were incarcerated.
6. Adding the Stable-2007 increased the accuracy of risk assessments in Canada but not in the USA (Alaska and Iowa).

For the Dynamic Supervision Project, the increased accuracy was small and the study has not been replicated or cross validated. Therefore, specific guidelines for using Stable-2007 to adjust risk based on Static-99 are premature. Already, though, the Static-99 is considered obsolete by its developers,[15] so it is hard to imagine how any risk assessor could directly apply the results of the 2007 Dynamic Supervision Project to a risk assessment in 2010. Even if that hurdle can somehow be overcome, there are at least two additional hurdles remaining for an SVP evaluator assessing an incarcerated person in the USA. This sole study was of offenders on probation (not incarcerated), and the combined risk-assessment procedure was essentially worthless for predicting sexual recidivism for the USA subjects (AUC = .60; 95% confidence interval = .42 to .76).

VII. CONCLUSIONS

Hanson (1998) provided valuable descriptions and proscriptions regarding adjusted actuarial evaluations, including: "The research can provide no easy answer as to when factors external to the actuarial instrument should be considered" (pp. 65-66).

With the exception of the "sexual-deviance" risk factor, data reported in meta-analyses (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005) have not been presented in a way that allows readers to compute accuracy statistics such as true positive and false positive outcomes. In turn, evaluators cannot use the data in those meta-analyses

[13] Checking with colleagues via professional Internet lists yielded no such studies. Thanks for input from Susan Sachsenmaier, personal communication on May 13, 2010.
[14] An AUC of .50 indicates no discriminative value. See Fan, Upadhye, and Worster (2006).
[15] "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." See "Static-99R Reporting Template: Routine Samples (Word or PDF)" at www.static99.org.

to identify the frequencies of false positives and false negatives associated with any risk factor.

When an evaluator uses the tables in the meta-analyses to consider whether a risk factor (e.g., empathy) not included in an actuarial instrument is actually associated with sexual re-offense, a lack of association (correlation or effect size near zero) would not support the use of that risk factor for adjusting the risk assessment based on the actuarial instrument. Unless multivariate analysis reveals interactive effects (e.g., moderator variables, suppressor variables), use of such variables to adjust risk would not be supported.

What about the factors in the meta-analyses that show significant correlation with sexual recidivism but are not included in an actuarial instrument? Hanson (1998, p. 66) wrote, "When the appropriate research has been conducted, what were considered external factors will either be incorporated into an actuarial instrument or will be deemed safe to disregard." As more data are collected, instrument developers can include additional items that add incremental validity, or change the weight or scoring of existing items.[16]

The meta-analyses (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005) are a good place to look for risk factors that are not associated with sexual recidivism, and a bad place to look for risk factors to use for adjusting actuarial risk assessment. Except for a relatively brief window of time between publication of a meta-analysis and revision (or not) of an actuarial risk-assessment tool, risk factors that help to predict sexual recidivism will already have been considered for use in actuarial tools. Those that add incremental validity to a risk assessment will have been incorporated into actuarial instruments. Those that have been considered and not incorporated in the instruments "will be deemed safe to disregard" Hanson (1998, p. 66).

In contrast to the factors listed in the meta-analyses, those listed as "Risk Factors outside the Static-99" at www.static99.org appear completely naked, with no empirical data at all. We encourage readers to look at that document and consider whether it can be used to answer "three questions that should be considered when evaluating an adjusted actuarial prediction" (Hanson, 1998, p. 67):

To what extent are the variables used for an adjustment already addressed by the actuarial instrument (i.e., how highly are they intercorrelated)?

Is there a compelling rationale or evidence strong enough to justify changing the actuarial prediction?

Is the amount of the adjustment proportional to the importance of the external variables being considered?

[16] Dynamic factors can be incorporated into an actuarial scale (e.g., the VRS-SO; Olver, 2004; Olver, Wong, & Nicholaichuk, 2008; Olver, Wong, Nicholaichuk, & Gordon, 2007), or developed into a separate scale (e.g., Stable-2007 and Acute-2007; Hanson, Harris, Scott, and Helmus, 2007). As with any risk-assessment tool, replication and cross-validation are necessary steps prior to clinical application.

To extend the metaphor further, we suggest that evaluators who would rely on "Risk Factors outside the Static-99" as a guide for adjusting an actuarial risk assessment would be flying blind with naked factors.[17]

Cautionary Note: Persons who intend to stand on the shoulders of giants should first peek beneath the giants' sandals.

In 1998, Hanson wrote that, given the state of knowledge at the time, he considered there to be three plausible approaches to sexual-recidivism risk assessment: guided clinical, pure actuarial, and adjusted actuarial. Eleven years later, Hanson and Morton-Bourgon (2009, p. 7) reviewed three studies that compared pure-actuarial to adjusted-actuarial sex-offender risk assessment, and concluded, "For all three measures, for all types of raters, and for all outcomes, the adjusted scores showed lower predictive accuracy than did the unadjusted actuarial scores." As we prepared an earlier draft of this paper, we accepted Hanson and Morton-Bourgon's conclusion uncritically, and we were ready to call for an indefinite moratorium on adjusted-actuarial sex-offender risk assessment. In the peer-review process for this article, a "blind" reviewer criticized our reliance on authority[18] and encouraged an independent review of the three studies (Gore, 2007; Hanson, 2007; and Vrana et al., 2008).

Our independent review does not lead us to concur with Hanson and Morton-Bourgon's interpretation of the data. We are troubled by our findings. As clinicians, we often rely on peer-reviewed research to guide our practice. Clinicians should be able to confidently take guidance from results and conclusions in a recent, relevant meta-analysis by leading researchers in the field. From the perspective of practicing clinicians (consumers of research), we beseech researchers to qualify their findings, consistently and carefully, especially in their concluding statements with important practical implications. We ask for strong, clear cautions regarding any summaries or conclusions drawn in part from studies that have not undergone peer review. We suggest that any data that are not (yet) publicly available should not (yet) be included in a meta-analysis. Alternatively, researchers could prominently display a warning such as "not ready for clinical application" for any meta-analysis that relies in part on data that have not been peer-reviewed and/or are not available for public scrutiny.

---

[17] Some of these factors are dynamic factors. Each and every one can change—and sometimes change substantially–over time. There is no guidance at www.static99.org whether these factors should be assessed in terms of an offender's current status, or in terms of his offense history. The unavailability of necessary guidance for this issue can only compromise levels of inter-rater reliability for these factors. Similarly, there is no guidance for how to assess these variables; when to assess them; how much, if any, unique variance the variables add to the Static-99 score; or how much weight to give these variables if adjusting the risk assessment based on the Static-99 score.

[18] "Things aren't 'true' just because Karl Hanson (or any other expert) says so. . . . There is an extensive quote concerning Hanson's evaluation of three studies that have compared actuarial scores and adjusted actuarial scores. Given the centrality of these studies to the basic point of this submission, the authors should review these studies on their own and provide their own assessment—not someone else's—concerning what was found."

In closing, we provide our own conclusion regarding the current state of knowledge. For the reader's convenience, we repeat the quote from Hanson and Morton-Bourgon (2009, p. 7):

> Three studies examined the difference between actuarial scores and adjusted actuarial risk ratings (Gore, 2007; Hanson, 2007; Vrana, Sroga, & Guzzo, 2008). In these studies, evaluators were required to complete an actuarial risk tool and then were allowed to adjust the final risk rating on the basis of factors external to the actuarial tool. All three studies were prospective, and evaluators completed the ratings as part of their routine procedures. In two studies, the raters were probation officers (Hanson, 2007; Vrana et al., 2008), and in the other study, the raters were either psychologists or correctional staff (Gore, 2007). For all three measures, for all types of raters, and for all outcomes, the adjusted scores showed lower predictive accuracy than did the unadjusted actuarial scores.

Our analysis of available data leads to a different summary regarding pure-actuarial versus adjusted-actuarial sexual-recidivism risk assessment: We are aware of no studies published in peer-reviewed journals that report empirical data directly comparing the accuracy of actuarial scores and adjusted-actuarial risk ratings. In her 2007 doctoral dissertation of assessments using the MnSOST-R, Gore (p. 61) reported that her analyses failed to reveal any statistically significant differences in overall accuracy between the pure-actuarial approach and the adjusted-actuarial approach. Additional research, including some projects in progress, may add to our understanding. We look forward to further developments in this field.

received January 20, 2010, revision submitted March 12, 2010, accepted May 12, 2010

REFERENCES

Bierman, K. L., Nix, R. L., Maples, J. J., & Murphy, S. A. (2006). Examining clinical judgment in an adaptive intervention design: The Fast Track Program. *Journal of Consulting and Clinical Psychology, 74,* 468-481.

Campbell, T. W. (2003). Sex offenders and actuarial risk assessments: Ethical considerations. *Behavioral Sciences and the Law, 21,* 269-279.

Campbell, T. W. (2007). Assessing sex offenders: *Problems and pitfalls-2nd Ed.* Springfield, IL: Charles C. Thomas.

Davis, D., & Follette, W. C. (2002). Rethinking the probative value of evidence: Base rates, intuitive profiling, and the "postdiction" of behavior. *Law and Human Behavior, 26,* 133-158.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668-1673.

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.

Doren, D. M. (2002). *Evaluating sex offenders: A manual for civil commitments and beyond.* Thousand Oaks, CA: Sage.

Epperson, D. L., Kaul, J. D., & Hesselton, D. (1998, September). Final report on the development of the Minnesota Sex Offender Screening Tool-Revised. Paper presented at the Annual Meeting of the Association for the Treatment of Sexual Abusers. Vancouver, BC, Canada.

Fan, J., Upadhye, S., & Worster, A. (2006). Understanding receiver operating characteristic (ROC) curves. Canadian Journal of Emergency Medicine, 8, 19-20.

Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice, 17,* 420-430.

Fitch, W. L., & Hammen, D. A. (2003). The new generation of sex offender commitment laws: Which states have them and how do they work? In B. J. Winick & J. Q. LaFond (Eds.) *Protecting Society from sexually dangerous offenders.* Washington, DC: American Psychological Association.

Friedrich, B. (2005). Correlates of sexual behavior in young children. *Journal of Child Custody, 2(3),* 41-55.

Gore, K. S. (2007). Adjusted actuarial assessment of sex offenders: The impact of clinical overrides on predictive accuracy. *Dissertation Abstracts International, 68*(07), 4824B. (UMI No. 3274898). Cited in R. K. Hanson & K. Morton-Bourgon. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21,* 1-21.

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

Hanson, R. K. (2002). Introduction to the Special Section on dynamic risk assessment with sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 14,* 99-101.

Hanson, R. K. (2003). Who is dangerous and when are they safe: Risk assessment with sexual offenders. In B. J. Winick & J. Q. LaFond (Eds.) *Protecting society from sexually dangerous offenders.* Washington, DC: American Psychological Association.

Hanson, R. K. (2007, March). *How should risk assessments for sexual offenders be conducted?* Paper presented at the Fourth Annual Forensic Psychiatry Conference, Victoria, British Columbia, Canada. Cited in R. K. Hanson & K. Morton-Bourgon. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21,* 1-21.

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

Hanson, R. K., & Morton-Bourgon, K. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73,* 1154-1163.

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21,* 1-21.

Hanson, R. K., Morton, K. E., & Harris, A. J. R. (2003). Sexual offender recidivism risk: What we know and what we need to know. In R. A. Prentky, E. S. Janus, & M. C. Seto (Eds.), *Sexually coercive behavior.* New York: New York Academy of Sciences (Vol. 989).

Hanson, R. K., & Harris, A. J. R. (2001). A structured approach to evaluating change among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 13,* 105-122.

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., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66,* 348-362.

Harris, G. T., & Rice, M. E. (2003). Actuarial assessment of risk among sex offenders. In R. A. Prentky, E. S. Janus, & M. C. Seto (Eds.), *Sexually coercive behavior.* New York: New York Academy of Sciences (Vol. 989).

Harris, G. T. & Rice, M. E. (2010). Assessment of risk and dangerousness in adults. In J. Brown & E. A. Campbell (Eds.) *Cambridge handbook of forensic psychology.* (pp. 299-306). New York: Cambridge University Press.

Harris, G. T., Rice, M. E., & Cormier, C.A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law and Human Behavior, 26,* 377-394.

Helmus, L., Hanson, R. K., & Thornton, D. (2009). Reporting Static-99 in light of new research on recidivism norms. *The Forum, 21(1), Winter 2009,* 38-45.

Helmus, L. (2009). *Re-norming Static-99 recidivism estimates: Exploring base rate variability across sex offender samples.* Unpublished M.A. thesis, Carleton University, Ottawa, Ontario, Canada.

Jaffe, E. (2004, May). What was I thinking?: Kahneman explains how intuition leads us astray. *APS Observer, 17(5),* 24-26.

Karouski, B. (February 24, 2010). Motor insurance premiums to double for motorists speeders. News-Insurances.com. Downloaded 3/2/2010 from http://www.news-insurances.com/motor-insurance-premiums-to-double-for-motorists-speeders/016713904.

Lilienfeld, S. O. (in press). Can psychology become a science? *Personality and Individual Differences* (2010), doi:10.1016/ j.paid.2010.01.024.

Murrie, D. C., Boccaccini, M. T., Turner, D., 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.

New Hampshire v. William Ploof (4 Feb 2009). Deposition of Amy Phenix, Ph.D. Docket No. 07-E-238.

New Hampshire v. William Ploof (4 Feb 2009). Deposition of Amy Phenix, Ph.D. Docket No. 07-E-238.

Nunes, K. L., & Cortoni, F. (2008). Dropout from sex-offender treatment and dimensions of risk of sexual recidivism. *Criminal Justice and Behavior, 35(1),* 24-33.

Olver, M. E. (2004). The development and validation of the Violence Risk Scale: Sexual Offender version (VRS:SO) and its relationship to psychopathy and treatment attrition. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 64(9-B),* 4628.

Olver, M. E., Wong, S. C. P., & Nicholaichuk, T. P. (2008). Outcome evaluation of a high-intensity inpatient sex offender treatment program. *Journal of Interpersonal Violence*, *24,* 522.

Olver, M. E., Wong, S. C. P., & Nicholaichuk, T. P., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale-Sexual Offender Version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19(3),* 318-329.

Petrila, J., & Otto, R.K. (2001). Admissibility of expert testimony in sexually violent predator proceedings. In A. Schlank (Ed.), *The sexual predator: Legal issues, clinical issues, special populations - Volume II.* Kingston, NJ: Civic Research Institute.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk.* Washington, DC: American Psychological Association.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk (Second Edition).* Washington, DC: American Psychological Association.

Rice, M. E., Harris, G. T., & Hilton, N. Z. (2010). The Violence Risk Appraisal Guide and Sex Offender Risk Appraisal Guide for violence risk assessment and the Ontario Domestic Assault Risk Assessment and Domestic Violence Risk Appraisal Guide for wife assault risk assessment. In R. Otto and K. Douglas (Eds.) *Handbook of violence risk assessment tools.* (pp. 99-120). Oxford, UK: Routledge/Taylor & Francis.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66,* 178-200.

Scalora, M. J. & Garbin, C. (2003). A multivariate analysis of sex offender recidivism. *International Journal of Offender Therapy and Comparative Criminology, 47,* 309-323.

Sjöstedt, G., & Långström, N. (2001). Actuarial assessment of sex offender recidivism risk: A cross validation of the RRASOR and Static-99 in Sweden. *Law and Human Behavior, 25,* 629-645.

Thornton, D., Beech, A., & Marshall, W. L. (2004). Pretreatment self-esteem and post treatment sexual recidivism. *International Journal of Offender Therapy and Comparative Criminology, 48,* 587-599.

Vrana, G. C., Sroga, M., & Guzzo, L. (2008). *Predictive validity of the LSI–OR among a sample of adult male sexual assaulters.* Unpublished manuscript, Nipissing University, North Bay, Ontario, Canada. Cited in R. K. Hanson & K. Morton-Bourgon. (2009). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73,* 1154-1163.

Wood, J. W. (1996). Weighing evidence in sexual abuse evaluations: An introduction to Bayes's Theorem. *Child Maltreatment, 1(1),* 25-36.