

Avoiding Garbage 2: assessment of risk for sexual violence after long-term treatment

BY GREGORY DeCLUE, PH.D., ABPP (FORENSIC)

The 1990s saw an upsurge in statutes and procedures for civil commitment of sexually violent predators. Some current cases involve consideration of whether a person who has been in long-term sex-offender treatment continues to meet commitment criteria. Psychologists' and psychiatrists' roles in such proceedings involve diagnosis, risk assessment, and risk communication. Particular challenges to evaluators are how to integrate and communicate findings regarding estimated risk from static and dynamic factors. Although both are theoretically important in considering a person's risk for sexual re-offense, there are considerably less empirical data regarding dynamic factors than static factors. Therefore evaluators should use considerable caution in using dynamic factors to adjust risk assessments based on static factors, and we should clearly communicate the lack of empirical base for risk-assessment adjustments based on dynamic factors.

Here we go again.

Over the past two decades, several states have instituted legal mechanisms for civil commitment of sexually violent predators. Florida did so five years ago. A little more than a year after that I had the opportunity to review a case in which three evaluators had made egregiously mangled risk assessments and testified in a manner that defied logic.¹ Those evaluations were some of the first of their kind done in Florida, and it showed. An early reviewer of that manuscript questioned the need for publication because the errors were so basic and obvious that she feared that readers would learn little from an article focusing on correcting them. But this was a real-life case with important consequences for the respondent and for potential future victims, and, unfortunately, the errors in the risk assessment were not unique to that case.

Now that Florida is nearly five years into civilly committing people who committed sex offenses, we are realizing that we need to figure out how to decide when to release them. The laws are in place, but evaluators are being called in to do something that is new to us. And once again we are making mistakes.

This article is sparked by a recent case in which eight evaluators performed risk assessments of one man's likelihood to commit new acts of sexual violence. Two evaluators were requested by the State to do evaluations in 2000 prior to the man's scheduled release from prison, and six evaluators were requested by the respondent to do evaluations prior to his civil commitment trial in 2004. Two of those six evaluators were called in late, never wrote reports, and did not testify. The other four evaluators requested by the respondent wrote reports, and three of them testified. Both evaluators requested by the State wrote reports and testified. This article focuses on the weaknesses—and occasional strengths—of those six reports and five testimonies.

Legal and ethical issues

In *Kansas v. Hendricks*, 521 U.S. 346 (1997), the United States Supreme Court ruled that it can be constitutional for states to civilly commit some sex offenders after they serve their criminal sentences. Several states have enacted laws and begun the civil commitment process, including a role for psychologists and psychiatrists to evaluate the respondent prior to a court hearing.

There is considerable scholarly debate about whether this type of civil commitment is fair or wise, and about whether and how scientist-practitioners can make accurate predictions.² Careful consideration of issues in this debate is important, but the debate does not preclude practitioners from conducting risk assessments now. After examining how courts have considered the issue, Janus and Meehl wrote, “[I]t seems well established that there is no constitutional impediment to using predictions of dangerousness in legal proceedings, up to and including those that may result in loss of liberty or death. As a legal matter, prediction is not, in all of its forms and for all purposes, so inaccurate as to violate the due process clause.”³

As mentioned above, the case discussed in this article arose in Florida. The following summary of the commitment statute is provided for context. Chapter 394 of Florida Statutes (2004) includes a section relating to civil commitment of sexually violent predators. The law is concerned with “a small but extremely dangerous number of sexually violent predators” who are “likely to engage in criminal, sexually violent behavior” (FS 394.910).

“Sexually violent predator” is defined as “any person who (a) has been convicted of a sexually violent offense; and (b) suffers from a mental abnormality or personality disorder that makes the person likely to engage in acts of sexual violence if not confined in a secure facility for long-term control, care,

and treatment” (FS 394.912(10)). The population affected by the law includes people who would otherwise be released from a prison or jail, a psychiatric hospital, or a juvenile commitment facility. “Likely to engage in acts of sexual violence” is defined as “the person’s propensity to commit acts of sexual violence is of such a degree as to pose a menace to the health and safety of others” (FS 394.912(4)).

Evaluators are given the task of gathering and analyzing data—including a direct examination of the person—to assist the trier of fact in determining whether the person meets criteria for civil commitment as a sexually violent predator. Either the Petitioner (the State) or the respondent can elect to have the case decided by a jury rather than the trial judge. The burden of proof is on the State, which must show by clear and convincing evidence that the person meets the criteria for being a sexually violent predator. If the person meets the criteria, he or she is confined indefinitely, with yearly reviews.

Three plausible approaches to risk assessment

It is not a new thing for psychologists and psychiatrists to be tasked with predicting which people who have committed sex offenses are likely to commit new offenses.⁴ In the past, such predictions were typically made on the basis of clinical interviews, observations in clinical settings, and some knowledge of the person’s history, but with little or no knowledge of research regarding quantifiable factors associated with increased or decreased risk for re-offending. This would now be called the *pure clinical approach* to risk assessment. Research suggests that clinicians using the pure clinical approach to risk assessment fair poorly, with risk estimates that are little better than chance.⁵

As research has been conducted and organized, there is a developing consensus that sexual recidivism is associated with at least two broad factors: enduring deviant sexual

interests and “antisocial orientation/lifestyle instability.”⁶ Sexual recidivism is higher for people whose self-report, offense history, and/or specialized testing reveals an enduring pattern of deviant sexual interests,⁷ presumably because they have urges to engage in deviant sexual behavior. Sexual recidivism is also higher for people who “are willing to hurt others to obtain their goals, can convince themselves that they are not harming their victims, or feel unable to stop themselves,”⁸ which may be manifested as or characterized by antisocial orientation, lifestyle instability, crime-prone personality, impulsive and/or reckless behavior, excessive drinking, frequent moves, fights, unsafe work practices, and a hostile and resentful attitude.⁹ Accuracy of risk assessments improves as these risk factors are taken into consideration.¹⁰

Because there is no single risk factor that is overwhelmingly more associated with re-offense than other risk factors, risk assessment accuracy is enhanced by combining risk factors. Three approaches to sexual re-offense risk assessment (other than the pure clinical approach mentioned above) have been identified: guided clinical, pure actuarial, and adjusted actuarial.

In the *guided-clinical* approach to risk assessment, an evaluator considers a range of empirically validated risk factors and then forms an opinion about the person’s risk for reoffense. In the guided-clinical approach, it is up to the evaluator to decide how much weight, if any, to give to the various risk factors in a given case. In the *pure-actuarial* approach, the evaluator considers a pre-determined set of risk factors and follows a pre-determined formula for weighing the combination of factors. In the *adjusted-actuarial* approach, the evaluator begins with an actuarial instrument but then may or may not adjust the estimated risk after considering additional variables not included in the actuarial instrument. Although at least one group of researchers has recommended the pure actuarial approach,¹¹ I know of no evaluators who

utilize a pure-actuarial approach in risk assessments prepared to assist judicial decision making.

Is there an empirical basis for favoring either an adjusted-actuarial approach or a guided-clinical approach over the other? I do not believe so. Most if not all studies that address the accuracy of actuarial risk-assessment instruments have measured the accuracy of the instrument itself, as if it were used in a pure-actuarial approach. There is little if any data to show that routine use of a procedure to adjust assessments based on additional factors enhances the accuracy of risk assessments.¹² Few studies have directly compared the accuracy of a guided-clinical approach versus a pure-actuarial or an adjusted-actuarial approach; one such study found no clear superiority.¹³ Hanson expects that research will lead to actuarial instruments that consistently yield more accurate risk assessments than those based on a guided-clinical approach, but that had not occurred by 1998.¹⁴ Nor has it occurred by 2004.¹⁵

Static and dynamic risk factors

Research in the 1990s often capitalized on readily-available data from files of people who had committed sex offenses, were later released, and had known records regarding whether they had been arrested and/or convicted of new sex offenses.¹⁶ Risk factors from these studies were simple, observable characteristics such as the number of prior sex offenses, victim characteristics (e.g., relative, known non-relative, or stranger; gender), and offender characteristics (e.g., age at time of release). Even a short list of such simple characteristics significantly enhanced the accuracy of risk prediction.¹⁷

The risk factors identified in the 1990s research tended to be static (fixed), such as history of childhood maladjustment or number of prior offenses. *Static risk factors* “mark long-term

propensities to engage in criminal behavior” but “cannot determine when offenses will occur, nor can they determine when offenders have substantially reduced their likelihood of re-offending (e.g., whether they benefited from treatment). For such assessments, dynamic risk factors are required. *Dynamic risk factors* are those that predict recidivism, have the potential of changing, and, when changed, are associated with corresponding increases or decreases in recidivism.”¹⁸ Dynamic risk factors can be further sub-divided into *stable* (relatively enduring) factors such as alcoholism and *acute* factors such as intoxication.

Risk assessments of people who have been in long-term sex-offender treatment programs should consider both static and dynamic risk factors. We will consider how to do so in a later section of this paper.

Sexual re-offense risk assessment instruments

In this case some evaluators used the SVR-20 (a guided-clinical sexual re-offense risk-assessment tool), Static-99 and/or MnSOST-R (actuarial sexual re-offense risk-assessment tools), and/or the PCL-R. Brief descriptions of those instruments and of the SONAR follow.

An evaluator using a guided-clinical approach utilizes an *a priori* list of risk factors such as the Sexual Violence Risk-20 (SVR-20). The manual includes: “The SVR-20 is an assessment method or procedure rather than a test or scale. Although it is an attempt to systematize the assessment of individuals, it is not sufficiently structured or standardized to be a test and does not yield norm-referenced or criterion-referenced scores.”¹⁹ The SVR-20 includes eleven items regarding Psychosocial Adjustment (e.g., sexual deviation, psychopathy, substance abuse problems), seven items regarding Sexual Offenses (e.g., high density, physical harm to victim(s), use of weapons or threats of death), and two items

regarding Future Plans (including attitude toward intervention).

The Static-99 is a risk-assessment instrument that utilizes only static (unchangeable) factors that have been found to correlate with sexual reconviction in adult males. "Static-99 is intended for males aged at least 18 years who are known to have committed at least one sex offense involving a child or a nonconsenting adult."²⁰ It is intended to be scored from the offender's official criminal record, but some information for some items can come from a clinical interview. There are ten items on the Static-99: prior sexual offenses, prior sentencing dates, any convictions for non-contact sex offences, current convictions for non-sexual violence, prior convictions for nonsexual violence, unrelated victims, stranger victims, male victims, age at release, and marital status. The coding rules were updated in 2003.²¹

The Minnesota Sex Offender Screening Tool-Revised (MnSOST-R) is a 16-item scale that was developed from three samples of male sex offenders (totaling 387 people) released from a Minnesota Correctional Facility.²² Scoring the items on the MnSOST-R requires accurate, detailed information about the number of sex convictions, length of sex offending history, force used in committing the sex offense, age of victim, relationship of the offender to the victim, substance abuse history, employment history, and other information.

The Hare Psychopathy Checklist-Revised (PCL-R) provides a reliable and valid means of measuring the clinical construct of psychopathy.²³ Psychopathy is generally considered as a personality disorder but has alternatively been considered (from an evolutionary psychology perspective) as an adaptive life strategy.²⁴ In either case, the features that define psychopathy can be placed into three broad categories: interpersonal, affective, and behavioral/lifestyle. "Interpersonally, psychopaths are grandiose, egocentric,

manipulative, dominant, forceful, exploitative, and cold-hearted. Affectively, they display shallow and labile emotions, are unable to form long-lasting bonds to people, principles, or goals, and are lacking in empathy and genuine guilt and remorse. Their lifestyle is impulsive, unstable, and sensation-seeking, they readily violate social norms and fail to fulfill social obligations and responsibilities, both explicit and implied.²⁵ Several studies have shown increased sexual recidivism for people with high PCL-R scores *and* evidence of sexual deviance as measured by penile plethysmograph (PPG)²⁶, a deviance rating scale,²⁷ or item 1 of the SVR-20.²⁸

Hanson and Harris have developed the Sex Offender Needs Assessment Rating (SONAR), which combines several different dynamic variables that have been associated with enhanced risk for sexual recidivism.²⁹ The instrument has coding rules for measuring intimacy deficits, social influences, attitudes tolerant of sexual assault, sexual self-regulation, general self-regulation, and acute risk factors *for people in the community on probation for sex offenses*. The instrument shows promise for enhancing risk predictions based on static factors, but I know of no data suggesting that assessment of such factors *during long-term confinement* assists in prediction of sexual recidivism *after release*.

Long-term treatment of sexually violent predators

As mentioned above, the United States Supreme Court ruled that it can be constitutional for states to civilly commit some sex offenders after they serve their criminal sentences. My reading of the case³⁰ is that a) the primary reason for civil commitment is to protect the public from dangerous people, and b) a state could confine someone even if it were expected that the person could not be successfully treated, but c) when a state does confine someone the state must provide treatment. The length of confinement/treatment will vary from person to person. In the nearly five years that

Florida has been civilly committing people, treatment center staff have not recommended release³¹ for any residents. At least two residents have died. Some may be recommended for release within the foreseeable future, since the clinical director testified in the case in question that treatment is envisioned to average about five to seven years.

Does long-term treatment of sexually violent predators reduce recidivism? At this point, no one knows. Some studies of recidivism after treatment have shown that people who complete sex-offender treatment have lower recidivism rates than people who were not treated, but due to methodological weaknesses it has been impossible to determine whether treatment has *caused* a decrease in recidivism risk. People who undergo some treatment and then either drop out or are “kicked out” might be expected to show some improvement, but not as much improvement as those who complete the whole treatment program. Not so! People who fail to complete treatment show *greater* re-offense rates than either people who completed treatment or people who had no treatment at all. Why? Some of the same people who are at increased risk to sexually re-offend may also be at increased risk to drop out of treatment because of lack of motivation, impulsiveness, or general belligerence.³² Thus treatment completion could be a marker variable for lower recidivism risk rather than a cause of it.

Consider a recent well-designed study. Hanson, Broom, and Stephenson report: “The treatment program examined in this study did not appear to be effective in reducing recidivism. Although some analyses slightly favored one group or the other, the differences between the treated and untreated groups was virtually zero after controlling for year of release, follow-up time, and static risk factors.³³ The data are fresh in that there was a 12-year follow-up of people released as recently as 1992, but the treatment delivered in the 1980’s is not the same as treatment being delivered now. One could argue that current treatment might be more effective at

reducing recidivism than treatment offered twenty years ago, and that the more extensive treatment offered in current long-term commitment programs might yield better results. But those are empirical questions and we do not yet have the data to answer them one way or the other.

Clinicians are designing long-term sex-offender treatment programs based on a consensus of treatment providers. But when one considers currently available research, we do not know whether sex-offender treatment works, we do not know what type of sex-offender treatment works (if any), and we do not know how to tell when (if ever) a person who has been at high risk to re-offend has made sufficient treatment gains that he is no longer at high risk to re-offend.

Some deficiencies in some risk assessments

This section presents my opinions about strengths and weaknesses in the reports and testimony in the case at hand. Generally, the evaluators requested by the Petitioner (the State of Florida) were strong in gathering and presenting information about static factors relevant to the respondent's risk for re-offense. Both of those evaluators paid attention to dynamic factors, including recent evidence regarding general self-regulation and sexual self-regulation, sexually deviant attitudes, cognitive distortions, and the respondent's relapse-prevention plan. However, compared to the evaluators requested by the respondent, both the evaluators requested by the State had less recent direct contact with the respondent and incomplete access to the most recent treatment records regarding the respondent.³⁴

The four evaluators requested by the respondent all mentioned or implied in their reports that they held the opinion that the respondent met criteria for civil commitment at one time, but no longer met the criteria. All four of these evaluators nominally used an adjusted-actuarial approach to

risk assessment, which was explicitly articulated as follows in one of their reports: “Currently, the best method for evaluating a sex offender’s recidivism risk is the clinically adjusted actuarial model.³⁵ Clinical adjustments to actuarial results are appropriate when research has demonstrated that information adds incrementally to the actuarial instruments’ predictive accuracy, and when the information is clearly beyond the static factors considered in that actuarial scheme.” Yet only one of these four evaluators (and not the one just quoted) presented data from the actuarial instruments in their reports, and in their reports and testimony not one of these four evaluators conveyed the extent to which there was empirical support for the additional variables they considered in adjusting from the high risk associated with the scores on the actuarial instruments.

Facts and opinions

Diagnoses,
risk
assessment
instruments,
treatment
progress

There was little disagreement among the evaluators regarding the data in this case. The two initial evaluators (both sent by the State) scored the risk assessment instruments in ways that led to scores associated with high risk, and subsequent, non-independent³⁶ scorings at the civil commitment center concurred. None of the four evaluators requested by the respondent disagreed significantly with the scoring of the risk assessment instruments or with the underlying data. A brief summary of the data regarding the respondent may be helpful to the reader at this point.

Mr. X was in his early 40s at the time of the trial. He had been arrested and charged with approximately 42 offenses, with four separate sex-crime cases, three of which led to conviction. His non-sex-related crimes were of varying types, including burglary, grand theft, drug crimes, fraud, major and minor traffic violations, a weapons charge, aggravated assault, and violations of probation. All evaluators agreed that Mr. X met criteria for diagnoses of Paraphilia Not Otherwise Specified

(NOS), Antisocial Personality Disorder, and substance-use disorders involving past abuse/dependence of alcohol, marijuana, cocaine, and hallucinogens. He had a history of mild-to-moderate depression and anxiety, but no history of psychosis or mania. On the basis of brief testing in prison, his IQ was estimated to be in the average to high-average range.

Various scorings of the Static-99 were all near 10, and various scorings of the MnSOST-R were around +16. The interpretation of these scores presented in the assessment at the civil commitment center follows: "On the Static-99 the resident received a Total Score of 11 [which] places him in the highest risk category for re-offense. ... The authors of the MnSOST-R suggest a cut score of +13 to identify the most dangerous sexual predators. Mr. X's score of +17 places him well beyond the cut point of the highest level of risk."

Also undisputed were the following facts: Mr. X completed his prison sentence around the beginning of 2001 and immediately began serving a sentence of 10-year probation. After a judge found probable cause, Mr. X waived speedy trial regarding the civil commitment, was confined at the civil commitment center, immediately signed up for treatment, and participated actively in treatment except when he was temporarily unable to do so due to medical problems. He had some set-backs in treatment due to violation of program and facility rules, but generally spoke and acted in a way that was interpreted to be progress in treatment. He completed some treatment modules successfully and was considered to be making progress in treatment at the time of trial. Although he was universally considered to be progressing in treatment, while at the treatment center he violated rules *and laws* to the extent that his probation could have been violated (but no violation of probation was formally reported).

Psychopathy
checklist

There was some difference of opinion about the scoring of the PCL-R. One of the initial evaluators (requested by the State) found Mr. X's score to be 35. The civil commitment

center found a score of 32. This level of difference on the PCL-R is not surprising. The standard error of measurement (SEM) is about 3.0 for single evaluations and about 2.0 for the average of two ratings.³⁷ Averaging these two ratings leads to a score of 33.5. Due to the statistical properties of the SEM, if 100 trained raters assessed the same subject at the same time, 68% of the scores would be expected to fall between 31.5 and 35.5, and 95% of the scores would be expected to fall between 29.5 and 37.5.³⁸

One of the subsequent evaluators testified that he scored Mr. X as 28 on the PCL-R, that he generally concurred with the previous scorings regarding Mr. X, but that Mr. X's progress in treatment contributed to the lowered score. Although this was not fully explored in direct or cross-examination, it appears that this evaluator did not follow standard procedure when scoring the PCL-R.

The PCL-R manual emphasizes this point by printing it in boldface: "Note that neither the PCL-R Total Score nor its Factor Scores can be used as a measure (e.g., pre-post scores) of treatment progress or outcome."³⁹ Why?

The PCL-R items are rated on the basis of the person's lifetime functioning as revealed by evaluations of the assessment data. Items should not be rated solely on the basis of present state or relatively recent behavioral history, each of which may be atypical of the individual's usual functioning because of extreme or unusual situational factors. ...

Novice raters are sometimes unsure about what to do if the individual's behavior is erratic or inconsistent, or if there was a dramatic change in his behavior at some point during the lifespan. ... [R]aters should score the PCL-R items according to the person's typical functioning; that is, on an evaluation of how he or she functioned, on the average, throughout the life span.⁴⁰

When scored according to the instructions in the manual, the PCL-R is a measure of the level of psychopathy a person has shown over his or her life span, not a measure of current state. Proper scoring of Mr. X should show little or no change

in PCL-R scores over the three years he has been in treatment because his not-so-psychopathic behavior during those three years should be combined with his very psychopathic behavior during the course of his life. It would, of course, be fine to communicate the results of the PCL-R with a notation that he has shown less psychopathic behavior in recent years *in his current setting*.

The
fundamental
attribution
error

How important is that last prepositional phrase: in his current setting? It can be very important. For example, in one study a group of psychiatrists predicted that fewer than one percent of U.S. citizens would engage in a particular behavior, yet in fact 90% of the citizens studied engaged in the behavior. "These experts on human behavior were *totally* wrong because they ignored the situational determinants of behavior in the procedural description of the experiment and overrelied on the dispositional perspective that comes from their professional training. Their error is a classic instance of the FAE at work."⁴¹

What is the FAE? This is the fundamental attribution error.⁴² "We are all subject to this dual bias of overutilizing dispositional analyses and underutilizing situational explanations when faced with ambiguous causal scenarios we want to understand. We succumb to this effect because our educational institutions, social and professional training programs, and societal agencies are all geared toward a focus on individual, dispositional orientations."⁴³

So what does the FAE have to do with Mr. X? At the times of the evaluations preceding his civil commitment trial, he said he was sorry for committing his sex offenses, that he now saw that they were morally as well as legally wrong, and that his actions victimized the underage males. He said he had never felt or thought that way before, but that treatment had changed his thoughts, feelings, and attitudes. He had developed a nice relapse-prevention plan, which he blithely showed to the evaluators. He also showed his sexual

autobiography (required as part of his treatment) in which he acknowledged in writing that he had committed sex crimes. Mr. X stated that he had engaged in numerous sexually deviant acts that were never documented via arrest or conviction. The four evaluators who had been requested by the respondent all testified that these changes reflected significant changes in Mr. X's disposition – that, in effect, he was not the same man who committed all those sex crimes. (For reasons unknown to me, some of those evaluators testified that Mr. X was still prone to commit *non-sexual* crimes but was not likely to commit new *sexual* crimes.)

I believe that it is important to consider possible situational explanations for Mr. X's change in behavior. Mr. X was in an unusual (for him) situation at the times of the evaluations preceding his civil commitment trial. He knew that the person to whom he was talking would be testifying at a trial that could result in him being confined, probably for years and potentially for the rest of his life. He had had the opportunity to read and study the evaluation reports of the two initial evaluators and to attempt to alter his speech and other behavior to create a more favorable impression regarding the risk factors identified therein. He did indeed talk and act differently than what would be expected on the basis of his previous behavior in the community and in prisons, but he is an intelligent psychopath and he knew he was in an evaluative situation.

Consider how much weight to give to dispositional or situational factors regarding a particular change in Mr. X's behavior. He was told that for treatment staff to recommend his release from the treatment program he would have to progress through all stages and phases of treatment, and to progress through a particular phase of treatment he would have to "pass" a polygraph examination. Prior to the polygraph examination he had not chosen to disclose any sexually deviant acts beyond those documented in arrest and court records. At the time of the polygraph he was told that

he was being “deceptive.” He then commented that he may have as many as 40 sex-offense victims, all underage males. It is of course possible that Mr. X underwent a deep psychic change that happened to occur coincidentally with being told that the polygraph indicated deception, but it seems much more likely that he recognized that he would need to change his verbal behavior if he were to achieve the goal of getting a favorable recommendation from treatment staff.

So did Mr. X change his verbal and nonverbal behavior because he is a changed man or because he was responding to situational determinants—or was it some combination of the two? I do not know, and I do not think that any of the evaluators in this case knew either. And that is what I think the evaluators should convey to the court.

Recommendations

If all six of the evaluators who assessed Mr. X were in general agreement about so many of the facts, why did the evaluators differ in their opinions about whether Mr. X was likely to engage in future acts of sexual violence? I believe that much of the difference arose as the evaluators shifted from empirically based analysis regarding static factors to subjective analysis regarding dynamic factors.

As evaluators communicate facts and opinions in our reports and testimony, it is essential to distinguish between our opinions and the factual bases for those opinions. Indeed, some would recommend against expressing an opinion about the ultimate issue of whether the person meets commitment criteria. Consider the role of analysts in an analogous task: “From their first workday, [Central Intelligence Agency] intelligence officers are told never to suggest policy. They are not policy makers and policy is not their concern.... [A]ll are made to understand their job is to present the best intelligence in a clear, concise, and unbiased manner.”⁴⁴ The

job of a CIA intelligence officer includes presentation of relevant data *and analysis of those data* to policy makers.

Similarly, the role of an evaluator does not stop at presentation of data, but rightly should include analysis:

In this role, psychiatrists [and psychologists] gather facts, through professional examination, interviews, and elsewhere, that they will share with the judge or jury; they analyze the information gathered and from it draw plausible conclusions about the defendant's mental condition, and about the effects of any disorder on behavior; and they offer opinions about how the defendant's mental condition might [affect] his behavior. ... Further, where permitted by evidentiary rules, psychiatrists [and psychologists] can translate a medical diagnosis into language that will assist the trier of fact, and therefore offer evidence in a form that has meaning for the task at hand. Through this process of investigation, interpretation, and testimony, psychiatrists [and psychologists] ideally assist lay jurors, who generally have no training in psychiatric matters, to make a sensible and educated determination about the mental condition of the defendant."⁴⁵

In communicating facts and opinions, evaluators must show their work. Then the trier of fact (the judge or the jury) can see where the evaluators were in agreement, where they disagreed, and why they disagreed. Although this was not consistently explained in their reports and testimony, I believe that the key area of disagreement was in how much weight to give to static versus dynamic risk factors.

We return to a brief quote from one of the evaluators' reports: In the adjusted-actuarial approach to risk assessment, "Clinical adjustments to actuarial results are appropriate when research has demonstrated that information adds incrementally to the actuarial instruments' predictive accuracy, and when the information is clearly beyond the static factors considered in that actuarial scheme." I expect that few evaluators would disagree with that in principle, but what happens when we consider that literally? *Research* does not include clinical lore, hunches, or "in my experience." "Research has *demonstrated*" would exclude pilot studies,

case studies, non-significant trends, or a single study that has not been cross-validated on an independent population. “[A]dds incrementally to the actuarial instruments’ predictive accuracy” means that there has to be scientific proof that the information from the additional variable must *add incrementally* to the accuracy of the risk prediction instrument. Show me a peer-reviewed journal article that shows cross-validated evidence that certain dynamic risk factors add incrementally to the accuracy of actuarial risk assessment and I will show you evaluators who are likely to agree about when it is reasonably safe to release a person from long-term sex-offender treatment. Problem: no such research. Result: disagreement.

In this case the pattern of agreement and disagreement among the evaluators mirrored what we know about sexual reoffense risk assessment: we are better at predicting who is at high risk to re-offend than we are at recognizing when—if ever—a person who was at high risk to re-offend has changed sufficiently that his risk should be considered to be moderate or low. Faced with this uncertainty, what can guide the best practice in risk assessment and in risk communication?

Risk assessment To my knowledge the best single source for understanding the relative accuracy of risk factors – and combinations of risk factors – for sexual re-offense is found in the 2004 meta-analysis presented by Hanson and Morton-Bourgon.⁴⁶ A meta-analysis is a research study that analyzes the results of previous studies. This particular meta-analysis examined 95 previous studies involving more than 31,000 sex offenders and close to 2,000 recidivism predictions. This research allows one to consider what is currently known about the strength of different risk predictors, using the standardized mean difference, d , as the effect size indicator.⁴⁷ Generally, the higher the d value, the greater the difference between recidivists and non-recidivists; “ d values of .20 are considered ‘small,’ values of .50 are considered ‘medium,’ and values of .80 are considered ‘large.’ The value of d is approximately

twice as large as the correlation coefficient calculated from the same data.”⁴⁸ A *d* value of 0 would mean there was no difference between recidivists and non-recidivists. For our purposes, the key findings are that the mean effect sizes for risk assessment instruments are relatively high (above the ‘medium’ range) and the mean effect sizes for dynamic variables associated with treatment success are relatively low (often *below* the ‘small’ range). For example, the mean effect sizes for the Static-99, MnSOST-R, and SVR-20 are .63, .66, and .77, respectively. The mean effect sizes for lack of victim empathy, denial of sexual crime, minimizing culpability, low motivation for treatment at intake, and poor prognosis in treatment (post) are -.08, .02, .06, -.08, and .14, respectively.⁴⁹

Now consider how this information could guide risk assessment of someone like Mr. X. Throughout his life he has shown many of the problem behaviors associated with a high risk for sexual re-offense, and as a result of that he scores very high on sexual re-offense risk instruments like the Static-99, MnSOST-R, and SVR-20. In the last few years he has been active in treatment and he shows apparent treatment gains in how he talks about victims, takes responsibility for his crimes, etc. It makes sense to begin risk assessment on the basis of an instrument with a medium or high-medium effect size (Static-99, MnSOST-R, SVR-20), but I do not see an empirical basis for adjusting that prediction based on variables with tiny effect sizes.

Risk
communication

I believe that evaluators should address treatment issues in our reports and testimony, but we should not pretend that what we know about the importance of dynamic risk factors is comparable to what we know about the importance of static risk factors. We should not communicate (directly or by implication) that we know more than we do about what—if anything—in sex-offender treatment produces consistent, lasting decreases in risk to re-offend. Reports and testimony about sexual reoffense risk should not move seamlessly from relatively well-established empirical relationships to speculation.

I will attempt to illustrate how evaluators could communicate information that might be useful even though there is little empirical data to support it.

In my experience as an expert witness in cases involving the civil commitment of sexually violent predators, *it appears to me* that when a respondent appears to be making progress in treatment, his⁵⁰ attorney is likely to emphasize that progress in presenting testimony to the judge or jury. Meanwhile, *it appears to me* that some assistant state attorneys are reluctant to elicit testimony about the unknown efficacy of treatment. *One assistant state attorney* recently commented that she did not want to open the door to testimony about the possibility that the respondent might never be released, and she stated that she preferred to present the case as if it were about getting help for the respondent, even though she believes, as I do, that the primary intent of the law is to protect the public. *I know of no data* as to whether most assistant state attorneys proceed similarly, but *if* that is a widespread attitude about how to present such cases, that *might* make it a bit more difficult for evaluators to convey the uncertainty of their findings during testimony. *If so*, the very distinction I am calling for evaluators to emphasize—between empirically based testimony about risk based on static factors and speculation about risk based on dynamic factors—may be actively avoided by both attorneys as the expert testifies.

Now consider, in abbreviated form, how a similar tone could be used in testimony about a person's risk to re-offend:

Mr. X clearly meets criteria for diagnoses of Paraphilia NOS, Antisocial Personality Disorder, and several substance-abuse/dependence disorders. Throughout most of his life he has engaged in a pattern of behavior that is similar to that of people who tend to have high recidivism rates. As a result, he scores high on sex-offender risk-assessment instruments that are derived from static factors. His lifetime pattern of behavior meets criteria for psychopathy, and both his pattern

of convictions and his self-report show an established pattern of sexual deviance. Research has shown that groups of offenders with both sexual deviance and psychopathy have considerably higher sexual re-offense rates than those with neither of those patterns. In sum, Mr. X's overall risk for future acts of sexual violence is considered to be high.

The effects of treatment on sexual recidivism are not well known. It is possible that treatment has no measurable effect on most people who have committed sex offenses. It is also possible that people who learn what is taught in modern sex-offender treatment will have the necessary tools to manage their deviant impulses, and that those who subsequently maintain life stability and avoid high-risk situations will be able to utilize what they have learned in treatment to decrease their risk of re-offense. If so, then it would be important to note that Mr. X says and does the kinds of things that treatment providers like to hear and see from people in sex-offender treatment. Although Mr. X acknowledges that he is still sexually aroused by underage males, he reports that he is learning how to manage his deviant sexual impulses.

Upon release, Mr. X would have several years of sex-offender probation. He would be required to continue in sex-offender treatment on an outpatient basis, and he would be monitored more closely than the average person on probation. He would be subject to random searches of his home, his body (via drug testing), and to the extent possible his mind (via polygraph testing). It is important to recognize that Mr. X has never successfully completed probation before, and that he has committed past serious offenses – including aggravated (nonsexual) assault and a sexual assault initiated by grabbing an underage male off a bicycle – while on probation. If he continues to behave as he has in the treatment program, then his risk for re-offense would be estimated to be somewhat less than that based on the static variables alone. If, instead, he behaves on his next probation as he has on all his previous probations and paroles, then he is likely to re-offend in a

variety of ways, including new sexual and non-sexual crimes as well as violations of the conditions of probation.

Unfortunately, there is little scientific research to guide us as we move from analysis of stable factors to dynamic factors. Although he has historically been at high-risk to sexually re-offend, Mr. X looks good now. [The following is optional, except in situations where evaluators are required to address the ultimate issue.] Although I cannot point to research to support my opinion, I do/do not consider Mr. X to be at high risk to re-offend because ...

Postscript

Some readers may recall that risk assessment of humans' propensity to violence has been compared to meteorological predictions about violent storms, leading to the conclusion that "Understanding how best to communicate assessment of risk is as important to mental health law as improving the validity of those assessments themselves."⁵¹ I was reminded of that article and of Mr. X as, while mulling over concepts for this article, I heard the following exchange (wording approximate):

National Public Radio (NPR) meteorologist John Hamilton: Hurricane Frances has been downgraded.

NPR host Scott Simon: Do we even call it a hurricane anymore?

Hamilton: Well, the winds have decreased from 130 miles per hour to 105 miles per hour, but that's still a category 2 hurricane.

Simon: So it's not just a breezy day in the park?

Hamilton: No, this is a powerful, potentially destructive hurricane.

Notes

1. DeClue, G. (2002). Avoiding garbage in sex offender reoffense risk prediction: A case study. *Journal of Threat Assessment*, 2, 73-92.
2. *Psychology, Public Policy, and Law*, 1998, 4 (1/2), entire issue. Campbell, T. W. (2000). Sexual predator evaluations and phrenology: Considering issues of evidentiary reliability. *Behavioral Sciences & the Law*, 18, 111-130.
3. 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, p. 36.
4. *Psychology, Public Policy, and Law*, *supra* note 2. See also Campbell, T. (2004). *Assessing sex offenders: Problems and pitfalls*. Springfield, IL: Charles C. Thomas. See also Doren, D. M. (2002). *Evaluating sex offenders*. Thousand Oaks, CA: SAGE.
5. Hanson, R. K. & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66(2), 348-362. Hanson notes that across 10 studies clinical judgments concerning risk for sexual recidivism correlated only .10 with observed recidivism and notes, for comparison, that the single item "prior sexual offenses" correlates .19. See also Hanson, R. K. & Morton-Bourgon, K. Predictors of sexual recidivism: An updated meta-analysis, 2004-02, p. 1. Downloaded 9/2/04 from http://www.psepc-sppcc.gc.ca/publications/corrections/pdf/200402_e.pdf.
6. Hanson & Morton-Bourgon, *supra* note 5, p. 1.
7. *Ibid.* See also Hanson & Bussière, *supra* note 5.
8. Hanson, R. K. & Morton-Bourgon, *supra* Note 5, p.1.
9. *Ibid.*, p. 1.
10. *Ibid.*
11. Quinsey, V. L., Harris, G. T., Rice, M. E., Cormier, C. A. (1998). *Violent Offenders: Appraising and Managing Risk*. Washington, DC: American Psychological Association.
12. So why adjust the estimated risk from an actuarial measure? Consider an extreme example: if it is well documented that due to severe illness the person has less than six months to live and will be severely physically limited during that time, it makes little sense to stick with an actuarial estimate of risk based on research regarding 15-year re-offense rates. Also I suspect that courts would give little weight to a risk assessment that completely ignored specific factual information about a given case solely because it did not fit into the boxes in the evaluator's actuarial risk scale.

13. Dempster, R. J. (1998). *Prediction of sexually violent recidivism: A comparison of risk assessment instruments*. Unpublished Master's Thesis, Simon Fraser University.
14. Hanson, R. K. (1998). What do we know about sex offender risk assessment? *Psychology, Public Policy, and Law*, 4, 50-72.
15. In the recent meta-analysis, the mean effect size of the SVR-20 (a guided-clinical assessment tool) is at least as large as those of the actuarial tools. See Table 1, page 32, of Hanson & Morton-Bourgon, *supra* note 5.
16. Note that recidivism studies never directly measure who commits a new sex offense, but must rely on observable variables such as arrest and conviction. Some people may be incorrectly counted as recidivists because they were arrested or convicted for crimes they did not commit. Others may be incorrectly counted as nonrecidivists because their crimes were not detected.
17. Hanson, R. K. (1997). *The development of a brief actuarial scale for sexual offense recidivism*. (User Report No. 1997-04). Department of the Solicitor General: Ottawa, Ontario, Canada.
18. Hanson, *supra* note 14, p. 51.
19. Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C.D. (1997). *Manual for the Sexual Violence Risk-20*. Vancouver, BC: British Columbia Institute against Family Violence.
20. 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.
21. Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. *Static-99 Coding Rules Revised 2003*. Downloaded 9/2/04 from http://www.psepc-sppcc.gc.ca/publications/corrections/pdf/Static-99-coding-Rules_e.pdf.
22. Epperson, D. L., Kaul, J. D., & Huot, S., Goldman, R., Alexander, W. (December 2003). *Minnesota Sex Offender Screening Tool-Revised (MnSOST-R): Development, validation, and recommended risk level cut scores*. Downloaded 9/2/04 from <http://www.psychology.iastate.edu/faculty/epperson/TechUpdatePaper12-03.pdf>.
23. Hare, R. D. (2004). *Hare Psychopathy Checklist-Revised (PCL-R): 2nd Edition*. North Tonawanda, NY: Multi-Health Systems.
24. *Ibid.*, p. 5. See, e.g., Mealey, L. (1995). The sociobiology of sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences*, 18, 523-599. The "adaptive life strategy" view of

psychopathy is informed, interesting, and intriguing, but is apt to be misunderstood unless one has a good grasp of the evolutionary psychology context.

25. Hare, *supra* note 23, p. 5.
26. Rice, M. E., & Harris G. T. (1997). Cross-validation and extension of the Violence Risk Appraisal Guide for child molesters and rapists. *Law and Human Behavior*, 21, 231-241.
27. Olver, M., & Wong, S. (2003). *Psychopathy, sexual deviance, and recidivism*. Manuscript in preparation. Cited in Hare, *supra* note 23.
28. Hildebrand, Ruitter, and Vogel (in press). Cited in Hare, *supra* note 23.
29. Hanson, R. K., & Harris, A. *The Sex Offender Needs Assessment Rating (SONAR): A method for measuring change in risk levels. 2000-1*. Downloaded 9/5/04 from http://www.psepc-sppcc.gc.ca/publications/corrections/200001b_e.asp.
30. Kansas v. Hendricks, 521 U.S. 346 (1997).
31. Treatment center staff can recommend release. If/when they do, a judicial decision is required for a person to be released. In the absence of a recommendation from the treatment center, residents of the treatment center may petition the court annually for release.
32. Laws, D. R., Hudson, S. M., & Ward, T. (Ed). (2000). *Remaking relapse prevention with sex offenders: A sourcebook*. Thousand Oaks, CA: Sage. Reviewed by G. DeClue: *Journal of Psychiatry & Law*. 30(2), 2002, 285-292.
33. Hanson, R. K., Broom, I., & Stephenson, M. (2004). Evaluating community sex offender treatment programs: A 12-year follow-up of 724 offenders. *Canadian Journal of Behavioural Science*, 36(2), 87-96, p. 94.
34. There appear to be two reasons for this, both reflecting finances. Evaluators are independent practitioners and bill for services hourly. Although the respondent is legally guaranteed the right to a speedy trial (within 30 days of the filing of the petition for civil commitment), some respondents waive that right and proceed to trial when, in consultation with their attorneys, they believe the time is right. In this case four years elapsed from the time of the initial evaluations to the time of the trial, with tentative trial dates set and then continued. Each time a potential trial date approached, attorneys had to consider the extent to which they wanted to authorize experts to spend time (and the state's money) updating the evaluations. The Assistant State Attorney attempted to balance the need for completeness with the need to hold costs to a reasonable level. The Assistant Public Defender (the attorney for the respondent) made little or no effort to limit costs.

35. No basis (either empirical or logical) for this claim was presented.
36. Results of the initial evaluators were known. Consistent with recognized best practice standards, treatment providers were not called to trial to present estimates of the respondent's risk to reoffend.
37. Hare, *supra* note 23, p. 67.
38. *Ibid*, p. 66.
39. Hare, *supra* note 23, p. 20.
40. *Ibid*, pp. 29-30.
41. Zimbardo, P. (2004). The psychology of evil. In A. G. Miller (Ed.), *The Social Psychology of Good and Evil*. New York, Guilford, pp. 21-50, p. 27. The behavior was ostensibly delivering a 450 volt shock to a hapless victim. The situational determinants included influence from an authority figure and observing a peer behaving obediently. For more details of the experiment see Milgrim, S. (1974). *Obedience to authority*. New York: Harper & Row.
42. Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), *Advances in experimental social psychology (Vol. 10, pp. 173-220)*. New York: Academic Press. Cited in Zimbardo, *ibid*.
43. Zimbardo, *supra* note 41, p. 24.
44. Anonymous (2004). *Imperial hubris: Why the West is losing the war on terror*. Dulles, Virginia: Brassey's.
45. Ake v. Oklahoma, 470 U.S. 68, 80 (1985).
46. Hanson & Morton-Bourgon, *supra* note 5.
47. $d = (M_1 - M_2) / S_w$, where M_1 is the mean of the deviant group, M_2 is the mean of the non-deviant group, and S_w is the pooled within standard deviation. In other words, d measures the average difference between the recidivists and the non-recidivists, and compares this difference to how much recidivists are different from other recidivists, and how much non-recidivists are different from other non-recidivists. See Hanson & Morton-Bourgon, *supra* note 5, pp. 6-7.
48. Hanson & Morton-Bourgon, *supra* note 5, p. 8, citing Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd Ed.)*. Hillsdale, NJ: Lawrence Erlbaum.
49. Hanson & Morton-Bourgon, *supra* note 5, Table 1, p. 32.

50. Almost all respondents are male.
51. Monahan, J., & Steadman, H. (1996). Violent storms and violent people: How meteorology can inform risk communication in mental health law. *American Psychologist*, 5 (9), 931-938, p. 937.