

[Home](#)

## What Community Supervision Officers Need to Know About Actuarial Risk Assessment and Clinical Judgment

[References](#)

*Patricia M. Harris*  
*College of Public Policy*  
*University of Texas at San Antonio*

[Research on Clinical Versus Actuarial Predictions of Offender Risk](#)  
[The Communication of Offender Risk](#)  
[The Importance of Clinical Judgment and Skill in Actuarial Assessment](#)  
[Conclusion](#)

**COMMUNITY SUPERVISION OFFICERS** as well as academics whose expertise includes offender classification may be surprised by the title of this article; after all, actuarial risk assessment tools have been available to the field of corrections in one form or another (e.g., Baird, Heinz, & Bemus, 1979; Hoffman, 1994; Nuffield, 1982) for over 25 years and reliance on actuarial risk prediction is now a fundamental precept of widely promoted evidence-based practices (Bogue, Campbell, Carey, Clawson et al., 2004). On the other hand, recent research on classification practices in community supervision indicates that support for actuarial risk assessment is far from universal. According to a national survey reported in 2001, roughly one-quarter of probation and parole agencies and approximately 44 percent of community corrections treatment providers contacted had not yet incorporated standardized risk assessment tools into supervision practice (Hubbard, Travis, & Latessa, 2001). In a later survey conducted by the National Institute of Corrections (2003) of 74 public community corrections agencies exercising jurisdiction over more than half of the probationers and parolees under supervision in the U.S., respondents expressed concerns about the accuracy of the instruments and their capacity to really measure offender risk.

Anyone who has worked with or studied probation and parole officers recognizes that skepticism about actuarial assessment, or alternately, belief in the supremacy of professional judgment about an offender's likelihood of new criminal activity, is prevalent in community corrections, even in agencies that have adopted "state of the art" tools. This article addresses three factors that may be interfering with greater investment by community supervision officers in actuarial risk assessment. The first is overconfidence in the influence of perfunctory discussions of the clinical versus actuarial debate. Due perhaps to the longstanding availability of statistical assessment tools in corrections practice, training in actuarial risk assessment emphasizes historical overviews of the development of actuarial prediction (i.e., the "generations" of assessment techniques), at the expense of more persuasive explanations of why or under what conditions so-called clinical or professional judgment yields less accurate predictions than empirically derived tools. A second potential impediment concerns shortcomings in risk communication and understanding; a growing body of research indicates that decision-makers' acceptance and utilization of assessment results may depend on precisely *how* offender risk is summarized. The third is the tendency to portray actuarial and clinical assessment methods as a black and white debate,

thereby failing to recognize and affirm the importance of clinical judgment and skill to the successful execution of actuarial risk assessment.

[back to top](#)

## **Research on Clinical Versus Actuarial Predictions of Offender Risk**

Studies that specifically address the efficacy of clinical judgment relative to actuarial assessment can be sorted into two categories. The first consists of research that compares the prediction accuracy of unstructured clinical judgments with that of actuarial tools. The second includes studies that compare structured clinical judgments with actuarial tools. Referred to simply as “clinical judgments” in most literature on this topic, unstructured clinical judgment involves the exercise of educated intuition, where information items gleaned from interviews, client history, psychometric instruments, and conferences with other professionals are engaged at the discretion of the individual carrying out the assessment (Meehl, 1954).

In contrast, structured clinical judgment refers to use of scores from formal instruments that either were designed and validated for purposes other than prediction of recidivism (and whose individual items may or may not exhibit correlations with recidivism outcomes), or that were developed to predict recidivism but which incorporate items selected for their plausibility (such as may be suggested by relevant literature) rather than as a result of statistical research during the creation of the instrument in question. An example of the former is the Hare’s Psychopathy Checklist-Revised (PCL-R), created solely for the purpose of measuring psychopathy, a clinical construct (Hare, 2003). Examples of the latter include the Historical-Clinical-Risk Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997) and the Multifactorial Assessment of Sex Offender Risk for Recidivism (MASORR; Barbaree, Seto, Langton, & Peacock, 2001). Douglas and Kropp (2002) prefer the label structured professional judgment over structured clinical judgment, in recognition of the numerous nonclinical professionals (such as probation officers and victim services personnel) who engage in risk prediction activities.

The key difference between actuarial and any clinical risk assessment is that the former includes only items known to correlate with outcome variables, as determined by statistical analysis of representative samples of cases followed up over fixed periods of time, such as three years. Depending upon the instrument, these items may also be weighted or otherwise subjected to mathematical manipulation and ultimately combined to optimize prediction accuracy. Higher scores on actuarial tools correspond to higher probabilities of reoffending.

### *Unstructured clinical judgment versus actuarial risk prediction*

Most conclusions about the inferiority of clinical predictions relative to actuarial ones are based upon research about unstructured clinical judgments. In the area of offender risk, the majority of these comparisons focus on the decision to release mentally disordered or violent offenders from psychiatric institutions, and their subsequent behaviors upon entry into the community. Researchers then assess which of the two methods, clinical or actuarial, better predicted both success and failure (e.g., reoffending) during a follow-up period. Examples of research producing results favorable to actuarial assessment include Quinsey and Maguire (1986) and Gardner, Lidz, Mulvey, and Shaw (1996).

Less common is research targeting more general populations of offenders, but again, analyses focus on success of institutional release decision-making. Wormith and Goldstone (1984), for example, examined the relative importance of subjective judgment variables (such as employment plans, prognosis upon release, and police recommendations) and static variables (such as criminal history, offense type, prior supervision outcomes, and demographic characteristics) in the statistical prediction of rearrest or revocation in a sample of 203 male offenders paroled from Canadian penitentiaries. The researchers found that inclusion of clinical variables in the model made only minimal improvements in the ability to correctly classify success and failures.

Influential descriptive reviews of compilations of research studies comparing the relative efficacy

of actuarial predictions and clinical judgments appear in Meehl (1954) and Grove and Meehl (1996). The latter work is particularly extensive, reporting on 136 published studies in a wide variety of contexts, including success in employment and education, adjustment to military life, psychotherapy outcome, and medical diagnoses as well as recidivism. Together, these reviews indicate that predictions based on clinical judgment only rarely outperform actuarial assessments and, more frequently, that the latter match or exceed clinical predictions in accuracy. Of the 136 studies examined by Grove and Meehl, 8 favored clinical judgment, 64 favored actuarial methods, and 64 exhibited “approximately equivalent accuracy” (p. 298).

These 136 studies were later subjected to a meta-analysis by Grove et al. (2000). The analysis of effect sizes revealed that statistical predictions outperformed clinical predictions on average 10 percent of the time, and greatly exceeded the latter in at least one-third of comparisons. The actuarial predictions of criminal or delinquent behavior reported in this study, by the way, were always more accurate than clinical predictions of the same.

The Grove and Meehl (2000) study is noteworthy not just for the number of studies analyzed but because the authors rule out competing explanations for the superior performance of actuarial predictions, including the assessor’s field of training, length of experience, and task-related experience. In addition, they confirmed that actuarial predictions were more accurate than clinical predictions, even when clinicians had access to and employed a greater number of variables than was available for use in the statistical prediction.

A different approach to comparing clinical and actuarial assessment appears in Mossman (1994), who calculated Area under the Curve (AUC) statistics for each of 44 published studies involving the prediction of violence in a total of 16,000 subjects consisting of parolees, psychiatric patients, and indictes. The AUC represents the probability that a subject has been correctly classified relative to a subject classified by chance (Hanley & McNeil, 1982). An AUC value of .50 indicates a prediction that does not improve upon chance. The higher the AUC value, the more accurate is the prediction and greater is the improvement upon chance. The AUC is now a favored statistic for summarizing prediction accuracy, not just because it takes both false positive and false negative errors into account, but because it is independent of both base rates (the frequency of negative outcomes in the sample in question) and cutoffs used for delineating high- from low-risk cases (Rice & Harris, 1995), which could otherwise favor prediction outcomes.

Unlike the works examined by Meehl (1954) and Grove and Meehl (1996), none of the studies included by Mossman reported direct comparisons of prediction methods, but rather outcomes from a single prediction method. Techniques represented in Mossman’s compilation included clinical judgment, use of past behavior as a prediction device, and discriminant analysis, both with and without cross-validation. By comparing the value of AUC across the various studies, Mossman demonstrated that clinicians were able to differentiate violent from nonviolent subjects with “a modest, better-than-chance level of accuracy” (p. 790), a finding that was unaffected by whether the prediction was short- or long-term. Predictions generated by discriminant analyses, including those involving cross-validation where shrinkage of prediction accuracy is the norm, were superior to those produced by other methods.

Studies favoring actuarial methods are not without their detractors. Litwack (2001) observes that efforts claiming to directly compare clinical and actuarial assessments really do not provide such contrasts; studies that employ offenders released from institutions as the population as subjects for follow-up do not include all offenders recommended for release and can include offenders released against the clinician’s judgment. Further, the clinicians whose judgments were studied may not have had access to the array of variables used in statistical predictions. Litwack also objects to the studies’ lack of cross-validation in both clinical and statistical predictions on new samples, which would allow for a comparison of shrinkage across the two methods.

### *Structured clinical judgment versus actuarial risk prediction*

Most contemporary research regarding the “clinical versus actuarial” question examines the relative efficacy of specific assessment tools (e.g., Barbaree, Seto, Langton, & Peacock, 2001;

Douglas, Yeomans, & Boer, 2005; Grann, Belfrage, & Tengstrom, 2000; Gray et al., 2004; Kroner & Mills, 2001). Characteristic of these analyses is the comparison of the accuracy of predicted outcomes obtained from the application of various structured clinical and actuarial instruments, using the same population of offenders.

Instruments of the structured clinical judgment type are like actuarial tools in that they are also founded in empirical research. They differ from actuarial tools in that items were not included specifically for their correlations with recidivism. Despite this difference, however, *overall* scores may predict outcomes of interest with fairly good success. This may be true even of instruments such as the PCL-R that were developed for purposes other than prediction of recidivism (Hemphill & Hare, 2004).

Several representative studies help to illustrate the kinds of outcomes produced by comparison of actuarial and structured clinical tools. Douglas, Yeomans, and Boer (2005) compared the predictive accuracy of five assessment tools using a randomly selected group of 188 male offenders released from federal institutions in Western Canada onto community supervision between 1989 and 1994, nearly all of whom (98.4 percent) had at least one conviction for a violent offense. Instruments used in the analysis included the Violent Offender Risk Assessment Scale (VORAS; Howells, Watt, Hall, & Baldwin, 1997), the Violence Risk Appraisal Guide (VRAG; Quinsey, Harris, Rice & Cormier, 2006), the Historical-Clinical-Risk Management-20 (HCR-20), the Hare's Psychopathy Checklist-Revised (PCL-R), and the Hare's Psychopathy Checklist: Screening Version (PCL:SV). Of the five tools studied, just two—the VRAG and VORAS—merited categorization as actuarial instruments.

The offenders were followed up for an average of 7.68 years. Analyses of the predictive accuracy of each instrument's total scores with respect to new violence revealed that the HCR-20 produced the highest AUC value. The value of the AUC for the HCR-20 total score was .82, compared with .79 for the VRAG, .76 for the PCL-R, .73 for the PCL:SV, and .61 for the VORAS. Particular components of some structured clinical instruments also produced high AUC values, including factor 2 of the PCL-R, with an AUC of .82, and the Risk Management scale of the HCR-20, with an AUC of .80. These results led the authors to conclude that a position of "strict actuarial authority" is unfounded.

Barbaree, Seto, Langton, & Peacock (2001) reported a comparison of the predictive accuracy of the PCL-R and six instruments commonly employed to assess sex offender risk, using 215 sex offenders released from a sex offender treatment program in a Canadian prison between 1989 and 1996 and followed up for an average of 4.5 years. In addition to the PCL-R, other instruments included the VRAG and five tools developed specifically for the prediction of sex offender risk: the MASORR, the Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice & Cormier, 2006), the Rapid Risk Assessment of Sex Offender Recidivism (RRASOR; Hanson, 1997), the Static-99 (Hanson & Thornton, 1999), and the Minnesota Sex Offender Screening Tool, Revised (MnSOST-R; Epperson, Kaul, & Hesselton, 1998). The VRAG, SORAG, RRASOR, Static-99, and MnSOSTR are actuarial tools; as noted earlier, both the PCL-R and MASORR are structured clinical judgment instruments.

The authors used three outcome measures: any recidivism, any serious recidivism, and any sexual offense recidivism. Analysis revealed that while some instruments were fairly good at predicting all three outcomes, no *one* instrument was superior in all three. For example, the SORAG yielded highest AUC values for any recidivism and any serious recidivism, at .76 and .73, respectively, but the RRASOR produced the highest AUC for any sexual offense recidivism, at .77. Of no small significance, the RRASOR is very easy to use and score. Finally, instruments falling within the structured clinical judgment category (i.e., the PCL-R and MASORR) yielded less successful predictions of new sexual offending than those that were actuarial.

Gray et al. (2004) examined the predictive accuracy of three instruments with varying clinical content in a sample of 315 mentally disordered offenders following release from a medium security institution in the United Kingdom between 1992 and 1999. Offenders were tracked for at least 3 years. Instruments included the PCL-R, the HCR-20, and the Offender Group

Reconviction Scale (OGRS; Copas & Marshall, 1998). Instruments were chosen for their differential emphases on clinical variables, with the PCL-R relying most heavily on clinical judgment, the OGRS using no subjective measures, and the HCR-20 falling in between. Using various analytic techniques, the authors consistently found that the OGRS yielded the most accurate predictions of reoffending.

In summary, comparisons of structured clinical judgment instruments with actuarial tools sometimes find that the former can produce results on par with or even better than the latter. More frequently, however, actuarial tools yield highest AUC values in prediction outcomes.

Comparisons like the ones summarized above have some limitations. Just because one tool happens to yield a higher AUC compared to another in the same sample of offenders does not mean that the difference between the instruments' AUCs is statistically significant. For example, Kroner and Mills (2001) conducted a comparison of the PCL-R, HCR-20, VRAG, Level of Service Inventory, Revised (LSI-R; Andrews & Bonta, 2001), and the Lifestyle Criminality Screening Form (LCSF; Walters, 1997). While actuarial instruments (VRAG and LSI-R) yielded more accurate predictions than the others, the differences between the AUC values were not statistically significant, suggesting a larger than desirable probability that they were due to chance factors in sampling alone. Hemphill and Hare (2004) point out that comparisons of the PCL-R with instruments designed to measure official indices of recidivism overlook its utility in predicting a wide range of antisocial conduct, such as institutional escapes, violations of community supervision, and deviant sexual interest. Thus, failure to include such behaviors in an outcome measure can suppress the value of the PCL-R in making predictions about offender risk.

*Which assessment method is better, and why?*

In summary, the case for or against clinical judgment is more complex than is typically represented to the field of community corrections. At this time, there is no strong empirical case to be made for risk assessments based on unstructured clinical judgments. Only rarely are their predictions of greater accuracy than actuarial methods; much more frequently, they are inferior.

Though some research demonstrates that *structured* clinical assessments can fare as well or better than some actuarial tools in some populations, there are important considerations that nonetheless tip the scales in favor of actuarial tools. Generally speaking, structured clinical tools are neither intended for nor have been validated on general populations of offenders. Certain instruments of the structured clinical variety can take longer to administer than the typical actuarial tool. For example, the PCL-R involves a three-hour interview. (While the PCL-R can also be coded from existing files, sufficient documentation would not be available in the files of most offenders entering community supervision, even if only felons were considered.) Whatever their superiority under some conditions and in specific populations, there is no compelling reason to favor these instruments over actuarial tools that have been purposely developed for use in assessing general populations of offenders under community supervision.

Readers not yet persuaded by this discussion should note that much corroborating research exists outside of the "actuarial versus clinical" arena to substantiate the suggestion that individuals acting on their own judgment are notoriously poor at estimating risk (see, e.g., Connolly, Arkes, & Hammonds, 2000; Kahneman, Slovic, & Tversky, 1982; Slovic, 2000a). Risk perception is an inherently error-prone process, affected by biases and speculative principles formed on the basis of a person's limited experience. For example, decision-makers typically fail to take base rates (the frequency with which the outcome of interest occurs in the population) into account when making predictions. To some extent this may be a function of not knowing base rates, yet research finds individuals will take base rates into account only when there is no other information and that they will give preference to irrelevant information, such as a stereotype, over base rates when both are available (Tversky & Kahneman, 1974).

Individuals generally overlook cumulative effects on risks when making predictions (Slovic, Fischhoff, Lichtenstein, 1982). Retrievability and salience factors also distort the accuracy of

subjective predictions. Retrieval bias refers to risk misrepresentation that occurs when easily remembered events are mistaken for frequently occurring events. Likewise, the salience (impact) of an event will cause decision-makers to unjustifiably heighten estimates of its reoccurrence (Tversky & Kahneman, 1974). Females and males tend to assign different probabilities to identical events, even when they have training in the same field (Barke, Jenkins-Smith, & Slovic, 1997). The interaction of the decision-maker's race and sex also accounts for variation in risk perceptions, possibly due to the perceived vulnerability of particular groups relative to others (Satterfield, Mertz, & Slovic, 2004; Slovic, 2000b).

In the absence of empirically based structured decision-making aids for predictions about offender risk, even the most trained and seasoned professionals make predictions that typically perform no better than chance (Monahan, 1981). Of course, some unassisted predictions can be less accurate than a chance-based classification system. Errors are especially likely when corrections practitioners are asked to make program decisions that have the potential to affect public safety. The tendency to make predictions that err on the side of public safety leads to unnecessarily restrictive decisions in offender release or supervision, otherwise known as overclassification. For example, Bonta and Motiuk (1990) compared rates of recommendations for placements in halfway houses across three jails, two of which employed the LSI-R as a classification instrument and one of which relied on a subjective decision-making tool. In the jails that used the LSI-R, 51 percent of assessed offenders were recommended for halfway house placement versus *only 16 percent* of offenders classified subjectively.

In short, actuarial tools are superior because they limit discretion that would otherwise result in erroneous and conservative predictions; take advantage of large quantities of information, as well as redundant and multiple measures, to maximize prediction accuracy; help to predict recidivism in diverse groups of offenders; and, in comparison to structured professional judgment tools, take less time to administer. On the other hand, the contributions of structured clinical judgment cannot be discounted with respect to certain populations of offenders. As Quinsey, Harris, Rice, and Cormier note (2006, p. 72), "human judgments applied in a very structured way play a large role in the actuarial prediction of violence."

[back to top](#)

### **The Communication of Offender Risk**

A growing body of research indicates that the manner in which a risk prediction is reported can alter the user's understanding and acceptance of assessment results. For example, Slovic, Monahan, and MacGregor (2000) found that forensic psychologists and psychiatrists were more likely to view a client's risk as higher when that risk was reported as a frequency (e.g., 10 out of 100 subjects with the client's characteristics are known to reoffend) than as a probability (e.g., the subject has a 10 percent likelihood of reoffending), though both portrayals clearly represent identical risks. Slovic and Monahan (1995) found that the range of the response scale provided to clinicians affected the magnitude of the risk decision they eventually rendered. That is, the clinicians were far more likely to assign lower probabilities to a client's future dangerousness when provided with a scale that had six values between 0 and 10 percent, than when provided with one that included just two values, 0 and 10 percent.

Two studies in particular help to highlight particular features of offender risk communication that may help to explain why some users remain resistant to actuarial instruments, however advanced empirical justifications for these instruments may be. While such studies focus specifically on the communication of risk of violence and employ only mental health professionals in the role of decision-makers, they have implications for communication of general risks of recidivism in community corrections.

#### *Resistance to quantification of offender risks*

In the first study, Hilton, Harris, Rawson, and Beach (2005) asked the question: What is the best way to "package" objective, statistical risk information to clinical staff working in a forensic

mental hospital to encourage more widespread use of that information? To answer these questions, the authors presented study participants with several variations of two hypothetical cases, the first, a lower-risk subject (.24 probability of recidivism over 10 years) and the second, a higher-risk subject (.64 probability over 10 years). Both were male patients in maximum security. The descriptions of hypothetical cases included either risk-relevant information (taken from the VRAG) or risk-irrelevant information (e.g., subject's weight, health, and personal preferences). Descriptions also included different summaries of subjects' likely risk, which took the form of one of the following a) a probability; b) a frequency; or c) a statement that no summary of risk was yet available. Taking all possible variations into account, there were 6 iterations of the high-risk case, and 6 of the low-risk case.

Next, the researchers asked the clinicians to 1) estimate the offender's likelihood of reoffending over the next 10 years, on a scale from 1 to 100; 2) rate the offender's risk compared to other forensic patients; and 3) report which information items most affected their assessment.

Results indicated that when given descriptive information about risk combined with the probability or frequency of reoffending, the clinicians tended to inflate their perceptions of the client's risk. That is, more risk information resulted in higher, but more inaccurate, estimates of actual risk. When clinicians were asked to rate likelihood of reoffending without having the benefit of either probability or frequency information, they overestimated the likelihood of reoffending for the low-risk case. Further, their estimates of the high-risk case were actually more accurate in the *absence* of risk-relevant information than when descriptive information was combined with percentages and frequency estimates. When asked to name the most important information items affecting their decisions, participants identified the subject's case history information as the most influential.

These results led Hilton et al. to conclude that though probabilities can be as persuasive as frequencies in conveying likelihood of reoffending, relevant case history information leads to inflation of offender risk estimates. Thus, decision-makers would be better off with only quantitative estimates of likelihood of re-offending, versus narrative information about a case.

This study indicates that the biases and heuristics known to distort perceptions of risk in persons acting only on subjective judgment may be present even when formal estimates of risk have been prescribed. In the field of community corrections, the officer who performs an assessment is typically the one who uses its results. This same individual would have access to and be aware of descriptions of prior offending, along with a great deal of other information disclosed by the offender during the assessment interview. In addition, the officer has opportunity to react to descriptive case details in the form of police reports and pre-sentence investigations. In combination, these facts may help to explain why some officers lack confidence in quantitative assessment results. Possibly, the content of training in risk assessment should be expanded to warn community supervision officers about the pitfalls of their attraction to descriptive information, and to the perils of emphasizing such inputs over assessment results.

#### *Preference for management-oriented risk communication*

In the second study, Heilbrun et al. (2004) sent a survey to a random sample of 1,000 psychologists, identified through the American Psychological Association membership database. The 256 psychologists who responded to the survey reviewed eight vignettes in which three variables appeared in diverse combinations. Variables included risk level (high, medium, or low); risk factors (static and/or dynamic versions of substance abuse, medication non-compliance, and violence); and risk model (prediction-oriented, involving a decision to civilly commit, where the court would relinquish jurisdiction over the offender upon commitment; or management-oriented, involving a decision to grant an inmate parole release, where the paroling authority would enforce conditions of supervision). For each vignette, respondents were asked to rate the relative usefulness of each of six different ways of communicating the subject's risk of committing a violent act toward others. Methods of risk communication included a) the probability of violence over a forthcoming period of months; b) age and status of the subject (such as whether the subject had a history of violence or substance abuse); c) subject's level of risk of committing a

violent act, stated as high, medium or low; d) a statement that the subject's risk of violence was dependent upon particular risk factors, with information about interventions that could control the risk; e) a statement that the subject was or was not dangerous; and f) likelihood that the subject would commit a violent act, stated as a percent. Of the six alternatives, d) represents a management model of risk communication, b), a descriptive model, and all others, predictive models.

Analysis of responses indicated that decision-makers least preferred "likelihood that the subject would commit a violent act" as a method for communicating risk, of the six alternatives. Second, decision-makers favored *prediction* models of risk communication in scenarios involving static factors, but preferred *management* models when dynamic factors were presented. This preference was further heightened in the presence of high risk.

In the context of community supervision, the study provides an additional explanation for officers' resistance to actuarial predictions. If they are like the clinicians who responded to Heilbrun et al.'s survey, officers find predictions inadequate when forecasts are not accompanied by identification of specific risk factors and appropriate interventions. According to Heilbrun et al., these enhancements may make decision-makers less likely to reject the prediction itself.

Development and implementation of a "fourth" generation of risk assessment tools such as discussed by Andrews, Bonta, and Wormith (2006), wherein outcomes are tied not just to predicted risk but the offender's needs, strengths, and responsivity factors as well, would go a long way toward addressing this potential hindrance to acceptance of actuarial tools. However, the findings of Heilbrun et al.'s study on communication of risk assessment suggest that these ties should be stated directly. Interestingly, some suggest that the deliberate pairing of an individualized statement of risk with an explanation of how risk factors may be modified to reduce risk is a means for mediating the "actuarial versus clinical polemic" that has permeated research on and practice of risk prediction (Webster, Hucker, & Bloom, 2002).

[back to top](#)

### **The Importance of Clinical Judgment and Skill in Actuarial Assessment**

A third potential stumbling block to greater acceptance of actuarial risk assessment is the choice attached, albeit indirectly, to the "clinical versus actuarial" debate. In embracing statistically derived assessment tools, officers may erroneously come to believe they must simultaneously relinquish carefully cultivated professional judgments and skills for determining offender risks. Regrettably, the phrase "clinical versus actuarial" is the source of a false dilemma, because successful execution of actuarial risk assessments is *utterly* dependent upon the officer's competent exercise of clinical judgments and skills.

To carry out a valid and reliable actuarial risk assessment, an officer must possess and draw upon effective interviewing techniques. These include a constellation of clinical skills that, when intertwined with apt professional judgments, promote offender disclosure, a prerequisite for accurate instrument scoring. Relevant skills include adequately reflecting back the feeling and meaning in offender's responses, to facilitate rapport but also to validate interviewer understanding of the subject's replies. Officers must be skilled in the art of the open-ended questioning technique, not just to maintain rapport but also to avoid limiting or otherwise influencing the content of the offender's response. In addition, officers need to be able to suspend judgment, affirm at times, and refrain from blaming and advising throughout the interview, for the sake of maximizing disclosure. Officers must recognize when they should ask for elaboration, such as when they hear information that contradicts earlier input, when a response is ambiguous, or when they simply have not received sufficient information to permit scoring of an item. To determine the meaning behind a lack of response, officers should also be able to recognize when the offender is confused, resistant, or merely anxious.

Some readers may recognize an overlap between effective assessment interviewing skills and both the general principles and opening strategies of Motivational Interviewing (Miller &



Rollnick, 2002). The officer's effective use of empathy, reflection, open-ended questions, and affirmation throughout the assessment interview begins a process of engagement with the offender that is gaining recognition as a means for increasing the latter's participation in treatment and reducing problematic behaviors (Moyers, Miller, & Hendrickson, 2005).

Skill must also be present to bridge multicultural divides. To conduct an effective assessment, actuarial or otherwise, officers must possess an adequate understanding of cultural differences as well as sufficient sensory acuity for determining whether and how those differences may impact client disclosure. How best to use eye contact (or whether to avoid it), how much time one should anticipate awaiting the offender's response, what tone and volume of speech to use, how close to sit to the offender, whether to expect narrative or direct replies, and how to word particular questions to prevent alienating the offender, are examples of knowledge officers can apply to the assessment context when interviewing individuals from cultures other than their own (see, e.g., Okun, Fried, & Okun, 1999; Severson & Duclos, 2005; Umbreit & Coates, 2000).

Officers must also draw upon considerable judgment and skill to score the assessment. Actuarial instruments can be comprised of both objective items that are easily scored as well as a variety of subjective items, such as questions regarding impulsivity, relationships, employment patterns, and attitudes that require extended consideration. Generally speaking, questions that are not clearly objective require the officer to reflect upon the offender's responses to multiple items in the assessment interview. Certain actuarial instruments even include items widely recognized as "clinical." For example, subject's overall score on the PCL-R, whether the subject meets DSM-III criteria for schizophrenia, and whether the subject meets DSM-III criteria for any personality disorder, all appear as items on both the VRAG and the SORAG (Quinsey, Harris, Rice, & Cormier, 2006). The PCL-R belongs to the category of structured clinical instruments, and assessment according to DSM criteria is quintessentially a clinical endeavor.

After scoring the instrument, officers must exercise professional judgment when determining whether an override of assessment results is appropriate. For instance, the *LSI-R User's Manual* (Andrews & Bonta, 2001, p. 12) cautions, "it is impossible to foresee all possibilities and assess all factors that may influence the likelihood of criminal behavior. The trained professional is encouraged to document features of an offender's situation that may require special consideration and that may even override the quantitative risk/needs assessment of the LSI-R."

Above all, officers must use judgment and insight following the assessment to determine appropriate referrals and fashion an effective supervision plan for the offender. Development of the supervision plan requires the officer to look back over the whole of the assessment interview to identify factors in the offender's current circumstances likely to aggravate the latter's continued involvement in crime, as well as forces that will help the offender to avoid it.

[back to top](#)

## **Conclusion**

This essay has addressed three potential stumbling blocks to wider acceptance and utilization of actuarial risk assessment in community corrections. While much research finds actuarial tools to be superior to alternatives, due either to better accuracy or a combination of accuracy and expediency considerations, the enterprise of actuarial risk assessment requires more comprehensive attention and must move beyond mere addition of new studies validating particular instruments, if a greater embrace by the community corrections community is to be achieved. This article has identified some areas where training about risk assessment may be improved. Future research should investigate which factors most affect officers' likelihood of accepting and acting on assessment results, including but not limited to scope of training in risk assessment and methods for communicating risk assessment results.

[back to top](#)

## **References**

Wilson, S.J., Lipsey, M. W., & Derzon, J.H. (2003). The effects of school-based intervention programs on aggressive behavior: A metaanalysis. *Journal of Consulting & Clinical Psychology*, 71, 136-149.

[back to top](#)

## **What Community Supervision Officers Need to Know About Actuarial Risk Assessment and Clinical Judgment**

Andrews, D. A., & Bonta, J. (2001). *The Level of Service Inventory-Revised*. Toronto: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency*, 52, 7-27.

Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, 28, 490-521.

Baird, S., Heinz, R., & Bemus, B. (1979). *Wisconsin case classification/staff deployment project: A two-year follow-up report*. Madison, WI: Department of Health and Social Services. Bureau of Community Corrections.

Barke, R., Jenkins-Smith, H., & Slovic, P. (1997). Risk perceptions of men and women scientists. *Social Science Quarterly*, 78, 167-176.

Bogue, B., Campbell, N., Carey, M., Clawson, E., Faust, D., Florio, K., Joplin, L., Keiser, G., Wasson, B., & Woodward, W. (2003). *Implementing evidence-based practice in community corrections: The principles of effective intervention*. Washington, DC: National Institute of Corrections.

Bonta, J. & Motiuk, L.L. (1990). Classification to halfway houses: A quasi-experimental evaluation. *Criminology*, 28, 497-506.

Connolly, T., Arkes, H. R., & Hammonds, K. R. (Eds.) (2000). *Judgment and decision-making* (Second ed.). Cambridge: Cambridge University Press.

Copas, J. & Marshall, P. (1998). The offender group prediction scale: A statistical reconviction score for use by probation officers. *Applied Statistics*, 47, 159-171.

Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior*, 29, 617-658.

Douglas, K. S., Yeomans, M., & Boer, D. P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior*, 32, 479-510.

Epperson, D. L., Kaul, J. D., & Hesselton, D. (1998). *Final report on the development of the Minnesota sex offender screening tool-revised (MnSOST-R)*. Paper presented at the 17 th Annual Conference of the Association for the Treatment of Sexual Abusers, Vancouver, Canada.

Gardner, W., Lidz, C. W., Mulvey, E. P., and Shaw, E. C. (1996). Clinical versus actuarial predictions of violence in patients with mental illnesses. *Journal of Consulting and Clinical Psychology*, 64, 602-609.

Grann, M., Belfrage, H., & Tengstrom, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behavior*, 27, 97-114.

- Gray, N. S., Snowden, R. J., MacCulloch, S., Phillips, H., Taylor, J., & MacCulloch, M. J. (2004). Relative efficacy of criminological, clinical, and personality measures of future risk of offending in mentally disordered offenders: A comparative study of HCR-20, PCL: SV, and OGRS. *Journal of Consulting and Clinical Psychology, 72*, 523-530.
- Grove, W. M. & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy and Law, 2*, 293-323.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30.
- Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29-36.
- Hanson, R.K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism* (User report 1997-04). Ottawa: Department of the Solicitor General of Canada.
- Hanson, R. K. & Thornton, D. (1999). *Static 99: Improving actuarial risk assessment for sex offenders* (User report 1998-01). Ottawa: Department of the Solicitor General of Canada.
- Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised (PCL-R)* (2nd ed.). Toronto, Canada: Multi-Health Systems.
- Heilbrun, K., O'Neill, M. L., Stevens, T. N., Strohmman, L. K., Bowman, Q., & Lo, Y.-W. (2004). Assessing normative approaches to communicating violence risk: A national survey of psychologists. *Behavioral Sciences & the Law, 22*, 187-196.
- Hemphill, J. F., & Hare, R. D. (2004). Some misconceptions about the Hare PCL-R and risk assessment: A reply to Gendreau, Goggin, and Smith. *Criminal Justice and Behavior, 31*, 203-243.
- Hilton, N. Z., T. Harris, G., Rawson, K., & Beach, C. A. (2005). Communicating violence risk information to forensic decision makers. *Criminal Justice and Behavior, 32*, 97.
- Hoffman, P. B. (1994). Twenty years of operational use of a risk prediction instrument: The United States Parole Commission's Salient Factor Score. *Journal of Criminal Justice, 22*, 477-494.
- Howells, K., Watt, B., Hall, G., & Baldwin, S. (1997). Developing programmes for violent offenders. *Legal and Criminological Psychology, 2*, 117-128.
- Hubbard, D. J., Travis, III, L. F., Latessa, E. J. (2001). *Case classification in community corrections: A national survey of the state of the art*. Final report of grant no. 98-IJ-CX-0008 awarded by the National Institute of Justice. Cincinnati: University of Cincinnati Center for Criminal Justice Research.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge : Cambridge University Press.
- Kroner, D. G., & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior, 28*, 471.
- Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law, 7*, 409-443.
- Meehl, P. E. (1954). *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence*. MN: University of Minnesota Press.
- Miller, W. R. & Rollnick, S. (2002). *Motivational interviewing* (Second ed.). New York: The

Guilford Press.

Monahan, J. (1981). *The Clinical Prediction of Violent Behavior*. Rockville, MD: National Institute of Mental Health.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783-792.

Moyers, T. B., Miller, W. R., & Hendrickson, S. M. L. (2005). How does motivational interviewing work? Therapist interpersonal skill predicts client involvement within motivational interviewing sessions. *Journal of Consulting and Clinical Psychology, 73*, 590-598.

National Institute of Corrections (2003). *Topics in Community Corrections: Annual Issue 2003*. Washington, DC: Author. Nuffield, J. (1982). *Parole decision-making in Canada*. Ottawa: Solicitor General Canada.

Okun, B. F., Fried, J., & Okun, M. L. (1999). *Understanding diversity: A learning-as-practice primer*. Pacific Grove, CA: Brooks/Cole.

nuffield, J. (1982). *Parole decision-making in Canada*. Ottawa: Solicitor General Canada.

Quinsey, V. L. & Maguire, A. (1986). Maximum security psychiatric patients: Actuarial and clinical prediction of dangerousness. *Journal of Interpersonal Violence, 1*, 143-171.

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

Rice, M. E. & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737-748.

Satterfield, T. A., Mertz, C. K., & Slovic, P. (2004). Discrimination, vulnerability, and justice in the face of risk. *Risk Analysis, 24*, 115-129.

Severson, M. and Duclos, C. W. (2005). *American Indian suicides in jail: Can risk screening be culturally sensitive?* Washington, DC : National Institute of Justice.

Slovic, P. (Ed.) (2000a). *The perception of risk* London : Earthscan Publications, Ltd.

Slovic, P. (2000b). Trust, emotion, sex, politics and science: Surveying the risk assessment battlefield. In P. Slovic (Ed.), *The perception of risk* (pp. 390-412). London: Earthscan Publications, Ltd.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Facts versus fears: Understanding perceived risk. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 463-489). Cambridge : Cambridge University Press.

Slovic, P., & Monahan, J. (1995). Probability, danger, and coercion: A study of risk perception and decision-making in mental health law. *Law and Human Behavior, 19*, 49-65.

Slovic, P., Monahan, J., & MacGregor, D. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior, 24*, 271-296.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Umbreit, M. S. and Coates, R. B. (2000). *Multicultural implications of restorative justice: Potential pitfalls and dangers*. Washington, DC: Office of Victims of Crime.

Walters, G. D. (1997). A confirmatory factor analysis of the Lifestyle Criminality Screening Form. *Criminal Justice and Behavior*, 24, 294-308.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart S. D. (1997). *HCR-20: Assessing risk for violence* (Version 2). Burnaby, Canada: Simon Fraser University, Mental Health, Law, and Policy Institute.

Webster, C. D., Hucker, S., J., & Bloom, H. (2002). Transcending the actuarial versus clinical polemic in assessing risk for violence. *Criminal Justice and Behavior*, 29, 659-665.

Wormith, S. J., & Goldstone, C. W. (1984). The clinical and statistical prediction of recidivism. *Criminal Justice and Behavior*, 11, 3-34.

[back to top](#)

## **Clinical Versus Actuarial Judgments in Criminal Justice Decisions: Should One replace the Other?**

Ægisdóttir, S., White, M.J., Spengler, P.M., Maugherman, A.S., Anderson, L.A., Cook, R.S., Nichols, C.N., Lampropoulos, G.K., Walker, B.S., Cohen, G., & Rush, J.D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *Counseling Psychologist*, 34 (3), 341-382.

Bunn, D., & Wright, G. (1991). Interaction of judgmental and statistical forecasting methods: Issues and analysis. *Management Science*, 37 (5), 501-518.

Bureau of Labor Statistics, U.S. Department of Labor, Occupational Outlook Handbook, 2006-07 Edition, Probation Officers and Correctional Treatment Specialists, on the internet at <http://www.bis.gov/oco/ocos265.htm> (visited on May 26, 2006).

Carroll, J.S. (1977). Judgments of recidivism risk: Conflicts between clinical strategies and base-rate information. *Law and Human Behavior*, 1 (2), 191-198.

Carroll, J.S., Wiener, R.L., Coates, D., Galegher, J., & Alibrio, J.J. (1982). Evaluation, diagnosis, and prediction in parole decision making. *Law and Society Review*, 17 (1), 199-228.

Darlington, R.B. (1986). Combining human judgment and multiple regression. *Journal of Personality Assessment*, 50 (3), 362-395.

Dawes, R.M. (1975). Case by case versus rule-generated procedures for the allocation of scarce resources. In M. Kaplan and S. Schwartz (eds.), *Human judgment and decision processes in applied setting*, pp. 83-94. New York: Academic Press.

Dawes, R.M. (1979). The robust beauty in decision making. *Psychological Bulletin*, 34 (7), 571-582.

Dawes, R.M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, (2), 95-106.

Dawes, R.M. (2005). The ethical implications of Paul Meehl's work on comparing clinical versus actuarial prediction methods. *Journal of Clinical Psychology*, 61 (10), 1245-1255.

Dawes, R.M. (1999). A message from psychologists to economists: Mere predictability doesn't matter like it should (without a good story appended to it). *Journal of Economic Behavior & Organization*, 39, 29-40.

Dawes, R. M., Faust, D., & Meehl, P.E. (1989). Clinical versus actuarial judgment. *Science, New Series*, 243 (4899), 1668-1674.

Douglas, K.S., Yeoman, M., & Boer, D.P. (2005). Comparative validity analysis of multiple