

Federal PROBATION

*a journal of correctional
philosophy and practice*

Volume 70 Number 2

Special Issue: Risk Assessment

[Introduction: Why Assessment "Matters" in an Evidence-Based Community Corrections System](#)

By James Byrne, Guest Editor

[Assessment with a Flair: Offender Accountability in Supervision Plans](#)

By Faye Taxman

[What Community Supervision Officers Need to Know About Actuarial Risk Assessment and Clinical Judgment](#)

By Patricia Harris

[Clinical Versus Actuarial Judgments in Criminal Justice Decisions: Should One replace the Other?](#)

By Stephen D. Gottfredson, Laura J. Moriarty

[Overcoming Sisyphus: Effective Prediction of Mental Health Disorders and Recidivism Among Delinquents](#)

By Albert Roberts, Kimberly Bender

[Mental Illness in Correctional Populations: The Use of Standardized Screening Tools for Further Evaluation or Treatment](#)

By Arthur Lurigio, James A. Swartz

[Risk Assessment and Sex Offender Community Supervision: A Context-Specific Framework](#)

By Andrew Harris

[Validating the Level of Service Inventory—Revised on a Sample of Federal Probationers](#)

By Anthony W. Flores, Christopher T. Lowenkamp, Paula Smith, Edward J. Latessa

[Validation of the Risk and Resiliency Assessment Tool for Juveniles in the Los Angeles County Probation System](#)

By Susan Turner, Terry Fain

[How much risk Can We Take? The Misuse of risk Assessment in Corrections](#)

By James Austin

[Assessing the Role of Clinical and Actuarial Risk Assessment in an Evidence-Based Community Corrections System: Issues to Consider](#)

By James Byrne, April Pattavina

[From Theory to Practice: The Lifecycle Document for the Results-Based Management Framework for the Federal Probation and Pretrial Services System](#)

[Contributors to This Issue](#)

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

Introduction: Why Assessment "Matters" in an Evidence-Based Community Corrections System

James Byrne, Guest Editor

Department of Criminal Justice and Criminology

University of Massachusetts, Lowell

Every day, federal, state, and local community corrections officers are required to make decisions about the "risk level" posed by individuals placed under community supervision. Based on this risk assessment, offenders are assigned to a specific supervision level, where a variety of "tools and techniques" will be applied in an attempt to "manage" the risk posed by these offenders, at least for their time under community supervision. With roughly 100,000 probation and parole officers nationwide supervising an offender population of close to 5 million, it can certainly be argued that risk assessment is the single most important decision made by probation and parole officers today. With large caseloads and shrinking budgets, "triage" is the "name of the game" in the community corrections field; and assessment is a critical first step in any triage process.

In this special issue of *Federal Probation*, a number of nationally known experts in the field of community corrections have been asked to provide their own assessments of this triage process, focusing specifically on the use of both clinical and actuarial risk assessments by community corrections officers. A number of questions about risk assessment are raised and answered in this issue, including:

- What do we actually mean by risk?
- How reliable and accurate are actuarial and clinical assessments?
- What types of assessment instruments are available for adult offenders, juvenile offenders, and specific offender groups, such as sex offenders, drug offenders, and/or mentally ill offenders?, and perhaps most importantly,
- What evidence exists that links initial (and ongoing) risk assessment (using actuarial and clinical assessment tools) with subsequent offender outcomes (i.e., the link between assessment and risk reduction)?

As the authors of the articles in this special issue demonstrate, we currently know more about how to *classify* offenders into several categories of "risk" of re-offending than we know about how to *reduce* their "risk" of re-offending while under community supervision. When viewed in this context, debates over the use of actuarial vs. clinical assessment tools tend to obscure a larger issue: Are we interested in short-term offender control or long-term offender change? I would argue that in an evidence-based community corrections system, a clear link needs to be established, beginning with 1) valid and reliable initial (and ongoing) assessment, continuing to 2) development of risk-specific community corrections interventions, and concluding with 3) identification of subsequent offender outcomes in communities with different *community* risk levels.

Because risk assessment is the “lynchpin” in this process, it is critical to the success of community corrections, both in terms of short-term offender control and long-term offender change. While much time and effort has focused on how to assess the risk level of *individual* offenders, far less research has been focused on the assessment of *community* risk level (see, e.g. Pattavina, et al., 2006). As we conduct further research on how different individuals (such as high- vs. low-risk offenders) respond to community supervision strategies in different communities (such as high- vs. low-risk neighborhoods), we will take the logical next step in the development of an evidence-based community corrections system.

In the first article of this special issue, “*Assessment with a Flair: Offender Accountability in Supervision Plans,*” Faye Taxman—one of this country’s leading experts on case planning in community corrections—argues that the first rule in evidence-based community corrections practice is that services (both treatment and control-based) to offenders should vary in intensity based on the risk level of offenders. She describes how the risk level of the offender should be considered: first in terms of the intensity and appropriateness of services, and next in terms of the offender’s individual case plan, which should focus on the clinically-based assessment of criminogenic needs (substance abuse, family dysfunction, peer associates, criminal personality, antisocial cognitions, low self control, and other factors, including mental health).

Patricia Harris expands on the notion that *both* actuarial and clinical assessments are essential components of an evidence-based community corrections system in her excellent review of both types of risk assessment in community corrections. Her article, “*What Community Supervision Officers Need to Know About Actuarial Risk Assessment and Clinical Judgment,*” reviews the empirical research on the adequacy of both clinical and actuarially-based risk assessment instruments, and then identifies three impediments to the full implementation of actuarial risk assessment in community corrections:

- unclear/cursory presentations on the purpose of actuarial risk assessments to line staff;
- poor communication of offender risk assessment results; and
- failure to recognize the importance of clinical judgment (and skill) in the (actuarial) risk assessment process.

The third article that focuses on the use of actuarial and clinical assessment tools—“*Clinical versus Actuarial Judgments in Criminal Justice Decisions: Should One Replace the Other?*”—is authored by two of the country’s foremost experts on risk assessment, Stephen Gottfredson and Laura Moriarty. They begin their review by offering the following unequivocal assessment of the two competing approaches to risk assessment: “In virtually all decision-making situations that have been studied, actuarially developed devices outperform human judgments.” While they argue that “there is a place for human judgment and experience in the decision-making process,” they offer the following caveat: over-reliance on human judgment may undermine the accuracy of the risk assessment, because probation and parole officers may “concentrate on information that is demonstrably *not* predictive of offender behavioral outcomes.”

While the three articles that have just been highlighted present an *overview* of existing research that has been conducted on both actuarial and clinical risk assessment instruments, the next three articles highlight the use of different types of assessment strategies for different types of offenders. First, Albert Roberts and Kimberly Bender examine the most commonly used risk and mental health needs assessment tools in juvenile corrections in their uniquely titled article, “*Overcoming Sisyphus: Effective Prediction of Mental Health Disorders and Recidivism among Delinquents.*” The authors discuss the implications of their review for the two central goals of youth assessment in juvenile justice settings: 1) the safety of the community, and 2) the rehabilitation (and clinical treatment) of individual juvenile offenders. [*Note: According to the “Myth of Sisyphus” webpage, “The gods had condemned Sisyphus to ceaselessly rolling a rock to the top of a mountain, whence the stone would fall back of its own weight. They had thought with some reason that there is no more dreadful punishment than futile and hopeless labor.”*]

Next, Lurigio and Swartz identify and critically review the use of actuarial and clinical risk assessment tools for the large number of adult offenders in our community corrections system

with serious mental illness (SMI). Focusing specifically on the instruments used to identify offenders with serious mental illness, Lurigio and Swartz describe the results of their recent research testing the accuracy of two “new generation” mental health screening tools in community corrections (the K-6 screening tool, and the Brief Jail Mental Health Screen).

Moving from mentally ill offender assessment devices to sex offender risk assessment devices, Andrew Harris provides a comprehensive review of the research on the use of actuarially-based and clinically-based assessments of sex offender risk. Harris’s article underscores one limitation—at least to some—of a *risk-driven* community corrections system: If we made decisions about sex offenders based solely on actuarial risk assessment, then very few sex offenders would ever be placed under close supervision, because as a group, sex offenders recidivate at remarkably low rates. The Harris article raises important questions for community corrections managers to consider, not only about the *accuracy* of sex offender-specific assessment devices using either actuarial or clinical assessment instruments, but also about the interplay between the actual *risk* posed by various types of sex offenders and the *stakes* (for offenders, victims, probation/parole officers, and the community) associated with decisions made on the appropriate level of control needed for this group of low-risk but high-stakes offenders. For sex offenders, over-classification appears to be an inevitable consequence of the “risk” assessment process, not because these offenders pose a high risk to recidivate, but rather because for this group of offenders, *any* risk—one in 10, one in 20 or even 1 in 100—is unacceptable due to the high stakes involved.

One of the challenges facing community corrections managers at the federal, state, and local level is the development of a *defensible* risk assessment system. By defensible, I am referring to a classification system that is externally reviewed and objectively validated. Two separate articles address this issue directly. First, Anthony Flores, Christopher Lowenkamp, Paula Smith, and Edward Latessa examine the predictive accuracy of the Level of Service Inventory-Revised (LSI-R) for a sample of 2,107 adult federal probationers, using subsequent incarceration (rather than the traditional re-arrest) as their outcome measure. Next, Susan Turner and Terry Fain present the results of their validation of the “Risk and Resiliency Checklist,” first developed in San Diego and currently being used in the Los Angeles Probation Department as the centerpiece of the department’s new automated case planning system. The authors found that a youth’s resiliency score—the net sum of risk factors (which have negative values) and protective factors (which have positive values)—is significantly related to subsequent recidivism, using re-arrest during a 12-month follow-up period (from initial assessment) as the outcome measure.

The final two articles in this special issue challenge much of the current thinking in the area of risk assessment and offer a different set of policy options for readers to consider. James Austin is one of this country’s foremost authorities on institutional and community-based classification systems. His article, “*How Much Risk Can We Take? The Mis-use of Risk Assessment in Corrections*” describes the six basic steps to follow in developing a risk assessment instrument:

- Risk assessment instruments must be tested on your (own) correctional population and separately normed for males and females,
- An inter-rater reliability test must be conducted,
- A validity test must be conducted,
- The instrument must allow for dynamic and static factors,
- The instrument must be compatible with the skill level of your staff, and
- The risk instrument must have face validity and transparency with staff, probationers, parolees, and policy makers.

Utilizing the concepts presented in his six-step risk assessment development model, Austin provides a critique of the best-known and most commonly used risk assessment instrument, the LSI-R, using illustrative examples from recent studies on the reliability and validity of the LSI-R that he conducted in Pennsylvania (parole board), Washington, and Vermont (community corrections). Austin concludes his review by identifying the “disconnect” that currently exists at the federal, state, and local level between an offender’s risk level and the availability of services: high-risk offenders are under-serviced, while low-risk offenders are over-serviced. As Austin suggests, “It would be helpful for those in the risk assessment business to start advocating a

more reasonable level of intervention that matches the risk they have so carefully calibrated” (this volume).

Byrne and Pattavina’s article, “*Clinical and Actuarial Risk Assessment in an Evidence-based Corrections System: Issues to Consider*,” concludes the special issue by identifying three topics central to the current debate over the use of actuarial and clinical risk assessment in community corrections: 1) the need to distinguish between risk assessment and risk reduction, 2) the dumming down of community corrections associated with the development of actuarial risk assessment instruments, and 3) the need to combine individual risk assessment and community risk assessment in the next generation of risk-driven supervision strategies.

Taken together, the ten articles included in this special issue of *Federal Probation* offer readers an opportunity to examine the empirical evidence on the application of actuarial and clinical assessment instruments in community corrections systems, and to consider the new wave of assessment instruments being developed for adult and juvenile offenders and for specific subgroups of offenders (mentally ill offenders, substance abusers, sex offenders). Each author raises challenging issues for policy makers, practitioners, and the public to consider regarding the proper role of assessment in an evidence-based community corrections system.

What “predictions” can be offered about the direction of the field? In the very near future, I suspect that the current focus on the reliability and validity of *risk* classification systems will be supplanted by discussions of how to improve *treatment* classification systems in both institutional and community corrections. When this occurs, it will likely be the result of an emphasis on offender change—rather than short-term offender control—as the primary purpose of our community corrections system.

[back to top](#)

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

[Home](#)

Assessment with a Flair: Offender Accountability in Supervision Plans

[References](#)

Faye Taxman, Ph.D.

Virginia Commonwealth University

[The Standardized Risk and Needs Assessment Tool Dilemma](#)

[Clarifying the Concepts of Risk & Dynamic](#)

[Risk Factors: Actuarial](#)

[Substance Abuse](#)

[Family Dysfunctional](#)

[Peer Associates](#)

[Criminal Personality](#)

[Antisocial Cognitions \(Attitudes/Orientation/Thinking\)](#)

[Low Self Control](#)

[Mental Health, Self-Esteem, Low Educational Attainment, Employment & Other Factors](#)

[Applying Rule #1 in Correctional Agencies](#)

[Conclusions](#)

RULE NUMBER ONE in EBP (evidence-based practice) is that high-risk offenders should be placed into appropriate treatment services, and that low- and moderate-risk offenders should not receive the same intensity of services. (Note: The use of the term “services” here includes both treatment and control techniques.) While this may seem like a simple concept, it encompasses the following: 1) use actual risk assessments; 2) use dynamic criminogenic needs; 3) adopt responsivity or matching strategies to link offenders to services and controls; and 4) administer heterogeneous programs that address the myriad of offender issues. The goal is to combine all of these together as a supervision plan that identifies the goals and specifies expectations for the offender. These expectancies become the binding agreements that define the criteria for being successful on supervision. Stated simply, assessment is not just a stand-alone process; instead, it is a process that should lead to the goal of a supervision plan that is designed to change the behavior of the offender.

Assessment should be the beginning of the correctional process. Of course in contemporary criminal justice practice, it can occur at a number of points, including arrest and pretrial detention, sentencing, intake to probation/parole or prison, and so on. In other words, it can occur in numerous places, all with slightly different goals—at pretrial to determine risk of flight or danger to society, at sentencing to determine the appropriate punishment and/or placement, at prison to determine security levels, and at probation/parole to determine risk to the community. In all of these calculations, the goal of the assessment is to inform decisions about the degree of restrictions that an offender should be given based on the offender’s history and seriousness of the current offense. The assessment can also contribute to what is traditionally referred to as a “treatment plan,” or more specifically the corrective action plan to help the offender become a productive citizen and contributing member of society. As noted recently by Ed Latessa and his

colleagues (2002), corrections practice today seldom ties the assessment to a plan for the offender. Instead the plan for the offender is generally made based on judicial or parole board decisions.

The most frequent stumbling block is an understanding of the core elements that are embedded in EBP Rule #1, and how to apply these elements in practice. That is, with the tools that are available, often there is a misunderstanding of the concepts of risk and needs. Often the terms “static” and “dynamic” are inappropriately interchanged with risk and needs. Risk refers to the actuarial (or statistical) likelihood that an offender will have further criminal behavior. Dynamic refers to the dimensions of the person’s functionality that, if improved, can affect their involvement in criminal behavior. A clarification of these concepts is the main goal of this paper, with an eye on trying to clarify how best to use these concepts in correctional practice.

[back to top](#)

The Standardized Risk and Needs Assessment Tool Dilemma

Don Andrews and his colleagues (2006) recently provided a historical review of the concept of objective assessment tools for the criminal justice system. The review detailed the generational development of assessment in the various phases of the criminal justice system over the last 80 years: clinical assessments of offenders’ risk to the community with some emphasis on treatment planning, actuarial risk assessment to assess the likelihood of further criminal behavior, actuarial risk assessment combined with dynamic variables to better guide treatment planning, and actuarial risk assessment tools supplemented by problem-specific tools. The development of standardized tools for the field has accompanied various needs in the criminal justice system, including classification, treatment planning, release decisions (from prison, jail, or parole), and sentencing. Essentially, assessment tools have been developed and used for various purposes, which adds to the complication of how to use the tool(s). Some tools are designed merely to identify risk factors related to certain decisions, while others are designed to identify the factors or needs that, if altered, can improve offender outcomes.

As is true for other fields, and as other articles in this issue of *Federal Probation* note (see papers by Austin and Harris), a major point of discussion in the criminal justice field has been the value of standardized assessment tools compared to that of subjective assessments by counselors and other correctional staff. The preference for subjective assessment is a long-standing issue in the field (as well as in psychology, education, and other disciplines), since professionals feel confident in their decision-making skills, and do not want to succumb to a paper-pencil test. But as discussed by Harris (2006) in another article of this edition, research persists in demonstrating that standardized objective tools enhance decision making, besides providing institutional safeguards against discretionary, biased, or inappropriate decisions. The use of standardized tools minimizes the potential for bias to be introduced into the decision making process by such human factors as the staff person being influenced by the dress, mannerisms, and/or attitude of the offender, in addition to such obvious factors as gender, class, and race.

Of course, the clinical vs. objective tool debate is an overstatement of the relative advantage of an interview with the offender. Third and fourth generations of assessment tools are accompanied by an interview (clinical in nature with “clinical” referring to an interview to collect information from the offender in a manner conducive to assessing the offender’s risk and needs). The purpose of the interview is to gather key information on key domains and then use that information to evaluate the offender’s responses in comparison to the official record (e.g., arrest records, crime report, treatment history etc.). That is, risk assessment with dynamic factors or the latest generation of risk assessment accompanied by specialized tools (such as drug screeners, mental health screeners, etc.) requires a clinical interview to obtain and assess information from the subject. A good assessment process requires interviewing the offender, which allows the criminal justice professional to gather, collect, and evaluate the offender’s responses along with other information obtained in official records. And, as promulgated by Taxman and colleagues (2004) and Taxman (2004), an important part of the interview process is engaging the offender

in processing his/her own responses to the interview questions as part of a process of engaging the offender in becoming more accountable for his/her behavior.

Lowenkamp, Latessa, and Holsinger (2006) found that many offenders were not screened for actuarial risk before being placed in community correctional programs in Ohio, and that reductions in recidivism were noticeable for high-risk offenders in correctional programs that tended to be multi-dimensional and primarily served high-risk offenders. The authors developed an actuarial risk tool that focused only on the offender's static risk factors (prior arrests, prior incarceration, age at current arrest, employed at arrest, history of failure in community correctional programs, drug use history). In a series of articles and presentations, they have reported the same results for offenders placed in residential programs, intensive supervision programs, and other correctional programs in Ohio. Using a quasi-experimental design, the researchers illustrate that reductions in recidivism are possible by using standardized risk tools, which help to ensure that high-risk offenders receive the more structured services. Their research also illustrates how poor classification schemes can result in over-classifying offenders (i.e., placing low-risk offenders in inappropriate programs) and only serve to increase the recidivism of this group of offenders. Their research basically supports Rule #1 of EBP regarding the importance of actuarial risk tools. This is the recent addition to a long-standing support for this concept from individual studies and also, more importantly, from recent meta-analyses (see meta-analyses such as Andrews, Bonta, and Hoge, 1990, Andrews et al., 1990; Lipsey & Wilson, 1998; Gottfredson, Najaka, & Wilson, 2001; Wilson, Lipsey, & Derzon, 2003).

Research studies of late have shown that the field is struggling with how best to use the concepts of risk and needs in criminal justice decisions, and particularly on how best to integrate dynamic or need factors. A series of articles in the 2006 *Crime and Delinquency* (edited by me and Doug Marlowe) illustrate how this struggle occurs. Taxman and Thanner (2006) detail how a randomized trial to examine the efficacy of a seamless probation-treatment protocol was affected by the classification of offenders as drug-involved. Offenders were assessed using an actuarial risk tool in one stage of the experiment and then a clinical assessment was conducted to determine drug use. Using the standard DSM-IV criteria (Diagnostic Statistical Manual IV-TR), a clinician assessed the offenders to be drug abusers. (DSM-IV states the accepted criteria for abuse and dependency.) In this experiment, half of the offenders were classified as high risk and half as moderate risk. However, few of the offenders in either the high risk or the moderate risk categories could be classified as drug dependent by the DSM-IV criteria. (The intervention involved a cognitive behavior treatment that was geared for offenders with drug problems.) Study findings indicate that the seamless system had no impact overall, but analysis found that the seamless system had a positive effect on high-risk and drug-dependent (addicted or serious abuse problems) offenders. In this study, the clinician did not use a standardized tool to assess for a drug problem, which resulted in overclassifying offenders as drug users when in fact many would not have met that criteria if a standardized tool was used. Another article in this edition by DeMatteo, Festinger, and Marlowe (2006) found that in many drug courts numerous offenders are not drug dependent and had generally low-threshold drug use (they were nevertheless classified as drug offenders largely due to their involvement in the legal system, which is one of the criteria for being classified as an abuser). Yet, these offenders are asked to participate in a highly structured program and required to go to drug treatment services. Not surprisingly, the drug courts do not tend to demonstrate reductions in recidivism.

[back to top](#)

Clarifying the Concepts of Risk & Dynamic

Risk and needs vs. static and dynamic? The third and fourth generation tools combine two concepts into one instrument or protocol. The two concepts are: 1) that actuarial risk factors can be used to determine the degree to which the offender's history predicts that he/she is likely to be a risk in the community or in a prison setting (i.e., the past predicts the future notion); and 2) that needs or psycho-social factors that should be ameliorated or addressed can be identified to reduce the risk for further involvement in the criminal justice system. The combination of these two concepts into one instrument or a cascading model (using screeners to determine the need for

more in-depth inquiry into a problem area based on the results from the screener, such as fourth generation instruments include) evolved from the needs of the criminal justice system for better classification and treatment placement tools. The researchers constructed the Wisconsin Risk and Needs Tool to allocate service resources accordingly (much like a triage approach, where high-risk offenders would receive the scarce resources first to prevent harm). This resource allocation tool was constructed on a management-model premise.

The field has had a difficult time learning to use these tools in a manner that would facilitate the intended purpose—Rule #1. Again, the intended purpose is to isolate the criminal drivers while keeping in mind the actuarial risk. Criminal drivers refer to the people, places, or things that affect an individual's involvement in criminal behavior. This means that the current status of an individual in areas that may in the past have been a problem may not be as relevant as other areas. Since many of these instruments use dichotomous (yes/no) responses or three categories of responses (none, some, many), practitioners are often left wondering how to select the drivers from other precursors. And, since many behaviors are intertwined, such as co-occurring disorders of mental health and substance abuse problems, the practitioner needs to determine which factors should be addressed as part of the criminal justice system, and which factors may be important for the person to address in the greater scheme of his or her overall health and well-being, but do not necessarily need to be encompassed in the criminal justice system.

Many attitudes, values, and behaviors lie on a continuum of “no problem” to “severe problem” behavior. This is important to keep in mind, because most human beings exhibit certain negative traits, but it is the degree to which these traits influence the subject's involvement in negative behavior such as crime or drug use that concerns the criminal justice system. In this case, the negative influence is one that predisposes an individual to engage in certain acts, or hold certain values or attitudes that they tend to hold when engaging in behaviors that are covered by the criminal laws.

[back to top](#)

Risk Factors: Actuarial

In determining actuarial risk, we measure behaviors that predict negative outcomes (increased risk for criminal behavior) (e.g. the concept of predictive validity). The actuarial risk generally refers to demographic or historical factors (past behaviors) that affect the trajectory of an individual. For example, age of first arrest (or incarceration) is a predictor of further involvement with the criminal justice system, since the earlier an individual has been involved in the criminal justice (or juvenile justice) system, the greater the likelihood of future involvement. The actuarial concept in criminal justice is similar to that used in assessing risk factors for health insurance (e.g. family history, age of onset of a disorder, number of occurrences, etc.) or car insurance (e.g. prior driving history, speeding violations, etc.). As discussed by Gottfredson and Moriarty (2006), the statistical methods and methodology for developing these tools are sound. The emphasis is placed on criminal behavior, and the historical factors that predict the likelihood that an offender will continue criminal conduct.

The key question is the criterion variable or the behavior that is being predicted. In traditional criminal justice literature, the criterion variable is new criminal behavior (as measured by new arrests or reincarceration). Yet, many proxies that may be used in a risk assessment may not be direct measures of criminal behavior. Examples are substance use (except the tautology that use of illicit substances is a criminal act) or other victimless crimes (e.g. prostitution, etc.), technical violation for failures on probation and/or parole supervision, and so on. Clarifying this concept is important to differentiate whether the behavior being predicted is actual further criminal behavior. It should be noted that heightened law enforcement activities (arrests) in some geographic areas (which increases the odds for arrest) may influence certain variables. This is why some researchers are focused on certain classes of behavior (e.g. property crimes, violent crimes, etc.) that are less susceptible to the neighborhood context that an individual resides in.

[back to top](#)

Dynamic Factors: Criminal Drivers

The third and fourth generation assessment tools include questions about dynamic factors, or psycho-social needs that, if unaddressed, tend to increase the risk that the individual will commit criminal acts. That is, while many of these factors may be present in most human beings, it is the degree to which they influence an individual's daily functionality that determines the degree to which they affect the offender's behavior (criterion validity). The important component is that these need factors also predict the likelihood that an individual will become involved in criminal behavior due to the impact on the offender's current behavior. Researchers have found that certain domains are more likely to negatively impact an individual, whereas other domains that we might think, using common-sense, have the same impact (e.g., mental health status, low educational status, or underemployment) are not directly related to criminal conduct.

[back to top](#)

Substance Abuse

Frequently the statement is made that over 70 percent of offenders are drug involved. This statement derives from reports regarding how many in the offender population report some use of illicit substances during their lifetime (or the lifetime prevalence). This statement many not refer to current use or use that is associated with dysfunctional behavioral. Using a clearer definition, researchers have generally found that about 35 to 50 percent of offenders have substance abuse patterns that require drug treatment (about one third of males, about half of the females) (Belenko & Peugh, 2005; Taylor, Fitzgerald, Hunt, Reardon, & Brownstein, 2001). The drug-crime nexus literature is a complex web that does not illustrate any causality between drug use and other criminal behavior, except for heroin or crack addicts, where the literature is clearer cut. The alcohol-crime nexus also is convoluted (besides the tautology of alcohol consumption in public, etc.), and just like the drug-crime literature, the relationship between substance use and crime depends upon the nature of the use and situation.

[Table 1](#) illustrates the criteria for abuse and dependency accepted in the field (APA, 2004). The DSM-IV criteria differentiate between use and abuse, both of which are defined by the degree to which the use (abuse) affects the person's daily functions. The literature on drugs and crime is most clear cut about the impact of providing treatment services for drug-dependent heroin and crack addicts—providing treatment will reduce recidivism and substance abuse. Based on this literature, it is suggested that it is important to identify drug-dependent addicts and then place them in appropriate treatment services. The priority should be given to targeting high-need (i.e. dependent) substance abusers for appropriate services. It should also be noted that those involved in the career business (i.e. entrepreneurs or those that are involved in dealing, etc.) may be classified as abusers when in fact their criminal behavior is linked to the business, and not to the drug use. While many involved in the business of drug dealing are also “dabblers” or users of small quantities of substances, their overall use is generally not due to compulsive behavior but rather to opportunity structures.

[back to top](#)

Family Dysfunctional

Family disarray and histories are generally precursors to learned behaviors—some are negative such as drug use or criminal behavior. Within this context, people learn attitudes, values, and behaviors. Differences exist in how families affect the behavior of men and women based on the degree of dysfunction in the family. For men, the stress from the family is to be a contributor (financial and otherwise) or to play a major role in the family. For women, the stress from the family is to be a caretaker or to be subservient to males in their lives. To obtain the support that is needed from the family, the offender is susceptible to responding to the pressure through criminal behavior (or drug use). The issues regarding the family are complex, in that the household may allow and tolerate certain behaviors in the home, including substance use or criminal behavior. And, the family could have expectations that the offender feels unable to

meet.

[back to top](#)

Peer Associates

The other (and sometimes more influential) support mechanism that many rely upon (non-familial) generally consists of peers or associates. The risk factor is that the offender associates with others in a like situation, and this reinforces the criminal behavior. Over time, the offender essentially loses contacts with prosocial or non-criminal individuals. In other words, the offender fails to maintain the social support network that supports mainstream behaviors (prosocial). This is not just an issue of whether the offender is involved in a gang but rather whether the offender has any close associates that are not connected to criminal behavior. The question here is the degree to which the offender relies upon the peers that are involved in the criminal justice system and whether any of the associates are non-criminally involved.

[back to top](#)

Criminal Personality

Using the DSM-IV criteria, antisocial personality disorder (ASPD) and impulsive behaviors are part of the composite of personality disorders. According to the DSM-IV, approximately 3 percent of men and 1 percent of women have some form of antisocial personality. As shown in [Table 2](#), the antisocial personality disorder is characterized by a callous unconcern for the feelings of others, gross or persistent attitude of irresponsibility and disregard for social norms, rules, or obligations, incapacity to maintain enduring relationships, low tolerance for frustration and low tolerance for use of aggression or violence, incapacity to experience guilt or to profit from experience, or marked proneness to blame others for the behavior that the offender exhibits. This personality disorder differs from psychopathy, which is a more callous version of an ASPD, and some states have developed legal or judicial definitions of what constitutes a psychopath. In terms of the medical definition (according to the DSM-IV), a psychopath is defined as having no concern for the feelings of others and a complete disregard for social obligations. The psychopath is generally considered callous and incapable of forming lasting relationships; the psychopath lacks empathy, remorse, anxiety, or guilt and tends to be devoid of conscience. Psychopaths are the extreme criminal personality. A proper diagnosis requires clinical skills as well as standardized tools (see the Hare's Psychopath Checklist, Hare, 1990).

[back to top](#)

Antisocial Cognitions (Attitudes/Orientation/Thinking)

As distinct from the personality disorder is the attitudes and cognitions of offenders. Yochelson and Samenow (1976) in their seminal work identified 36 thinking errors that they believed are used to shun responsibility, at least as defined by society's standards. The continuum of criminality is from responsible to irresponsible; under "irresponsible" behavior, there is a range from nonarrestable to arrestable behavior. The scholars contend that all individuals have these kinds of thinking errors; however, criminals exhibit more of them, and they tend to be on the irresponsible end of the continuum in many of the areas. The thinking error phenomenon gained further steam with the work by Walters (1990) and his colleagues, in which they developed eight subscales to measure criminal thinking; Walters' work influenced a new tool on criminal thinking errors (Knight, Garner, Simpson, Morey, & Flynn, 2006). Most of this work has been criticized because the tools to measure criminal thinking have not been validated on a non-offender population, and therefore it is unclear whether these characteristics are concentrated in offenders or distributed among the general population. Also many of the "thinking errors" are common defense mechanisms that are used by human beings to handle situations.

The typical thinking errors include dominance, entitlement, self-justification, displacing blame, optimistic perceptions of realities, and "victim stance" (e.g. blaming society because they are considered outcasts). As noted by Mark Lipsey and Nana Landenberger (2006), such "distorted

thinking may misperceive benign situations as threats (e.g., be predisposed to perceive harmless remarks as disrespectful or deliberately provocative), demand instant gratification, and confuse wants with needs” (p. 57). The issue about attitudes and orientation is that the focus is on how the offender processes and interprets information.

[back to top](#)

Low Self Control

Impulsive and risk-taking behavior is another dynamic characteristic of offenders. The general premise is that low self control does not define criminal behavior; instead, it provides a context for criminal acts depending upon opportunities and other motivating factors. A person’s decision to engage in criminal acts is affected by other factors such as natural constraints, attachments to parents, school, employment, and so on (Gottfredson and Hirschi 1990, p. 95-97). Low self control is exhibited by the offender being easily persuaded by situational and environmental factors, and without attachments there is little to constrain the individual.

[back to top](#)

Mental Health, Self-Esteem, Low Educational Attainment, Employment & Other Factors

Mental health status, self-esteem, low educational attainment, low employment options, and other factors are frequently discussed in the realm of criminogenic needs. The definition of a criminogenic need is that the factor predicts criminal behavior, and the research literature does not demonstrate that the presence of these attributes predicts recidivism or involvement in criminal behavior. Rather, low educational attainment and unemployment appear to be correlated with the offender population, which leads some to conclude that addressing these factors may also reduce recidivism. As discussed previously with other substance abuse and ASPD, the behaviors range on a continuum. The same is true with mental health disorder, where the problems range from anxieties or depression to erratic and/or risky behavior (e.g. hears voices or expresses disorganized, disoriented, or paranoid thoughts; appears lethargic and sad; unusually manic in behavior, etc.). However, it is generally recognized that in order to improve an offender’s wellbeing (which may not be related to recidivism reduction efforts), he or she would benefit from improved employment, educational, and mental health status(es). Ultimately, addressing these issues may affect the ability of the individual to be a contributing member of society and/or family; it is unclear whether addressing these factors in and of themselves will affect criminal behavior.

[back to top](#)

Applying Rule #1 in Correctional Agencies

[Exhibit 1](#) illustrates the implementation of these principles into a model. Essentially, actuarial risk level should be determined to identify what is the offender’s likelihood of further criminal behavior. High-risk offenders should be targeted for treatment-based on the area (s) in which they score moderate or high on criminogenic needs. That is, the offender needs to be assessed also on the criminogenic needs to identify the drivers to their criminal behavior. The notion is that, similar to treatment placement models, actuarial risk should drive the priority for intensive control and appropriate services, with a focus on selecting programs that address multiple problem areas. “Appropriate” refers to attention to the criminogenic factors that have been identified.

The model presented in the exhibit illustrates how the criminogenic factors can exist regardless of risk level. That is, a substance abuser may be low risk due to the fact that he or she does not have a history in the criminal justice system. Other criminogenic factors may exist in that low-risk person, but they are more likely to be low to moderate in severity. As the offender moves along the continuum of risk (moderate to high), then it is more likely that more severe problem behaviors may occur. This is a byproduct of the offender’s inability to be a productive, contributing member of society. For example, a high-risk offender may have criminogenic needs

relating to self control, peer associates, ASPD, and substance abuse. The combined treatment and control strategies should be designed to address these issues. The model also suggests that the high-risk offender is more involved in situations, settings, and individuals that are likely to further their criminal conduct. Hence, control and treatment services should be concentrated on this individual to achieve the desired goal of reducing the risk of recidivism.

[back to top](#)

Conclusions

The purpose of this article is to further elaborate on Rule #1 in Evidence-based practices to better illustrate the concepts and to define criminogenic needs in the context of risk level. This article is driven by the needs of the field to translate the principle into operational terms. An actuarial-based risk screen is important to determine the degree to which offenders should be given services and resources to ameliorate criminal behavior. The type of services is determined by how the offender “scores” or presents on several criminogenic areas. Those offenders with high criminogenic needs, particularly those that are high or moderate risk, should be given services to ameliorate the criminogenic need, which should reduce the risk for recidivism. Exhibit 1 conceptually presents the framework underscoring EBP #1; the challenge to organizations is to implement this principle.

The field faces several challenges relating to organizational stamina in implementing Rule #1 by following this core concept. The first challenge is the willingness of the organization to focus services on high-risk offenders, which generally means that moderate-or low-risk offenders should not be given such services. Minimizing the provision of services for low-risk offenders essentially results in decisions that the probation supervision should minimize the disruption from prosocial behaviors, since they are likely the glue that is preventing the offender from becoming criminally involved. Another factor is that the case plan/supervision plan should be driven by the goal to ameliorate the criminogenic drivers. This is critical, since it provides the formula for reducing the risk of offenders in the community. Exhibit 1 illustrates that when an individual is identified as having moderate or high criminogenic needs, then the plan should be to address these criminal drivers in the case plan. That is, the results of the assessment are directly relevant to the components of the case plan, because it provides an avenue to assist the offender in attending to issues that are relevant to his or her life. In short, EBP #1 challenges the organizations to redo case plans so that they address the drivers (criminogenic needs) that are more pertinent to the situational factors of the offenders. In so doing, case plans become the glue for the offender that addresses the risk factors.

[back to top](#)

References

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

Table 1: DSMIV-TR Criteria for Substance Abuse for Dependency

Criteria for Substance Abuse

A pattern of substance use leading to significant impairment or distress, as manifested by one or more of the following during in the past 12 month period:

1. Failure to fulfill major role obligations at work, school, home such as repeated absences or poor work performance related to substance use; substance-related absences, suspensions, or expulsions from school; neglect of children or household
2. Frequent use of substances in situation in which it is physically hazardous (e.g., driving an automobile or operating a machine when impaired by substance use)
3. Frequent legal problems (e.g. arrests, disorderly conduct) for substance abuse
4. Continued use despite having persistent or recurrent social or interpersonal problems (e.g., arguments with spouse about consequences of intoxication, physical fights)

Criteria for Substance Dependence

Dependence or significant impairment or distress, as manifested by 3 or more of the following during a 12 month period:

1. Tolerance or markedly increased amounts of the substance to achieve intoxication or desired effect or markedly diminished effect with continued use of the same amount of substance
2. Withdrawal symptoms or the use of certain substances to avoid withdrawal symptoms
3. Use of a substance in larger amounts or over a longer period than was intended
4. persistent desire or unsuccessful efforts to cut down or control substance use
5. Involvement in chronic behavior to obtain the substance, use the substance, or recover from its effects
6. Reduction or abandonment of social, occupational or recreational activities because of substance use
7. Use of substances even though there is a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance

Source: ApA, 1994.

Table 2: DSMIV-TR Criteria for Antisocial Personality Disorder (ASPD)

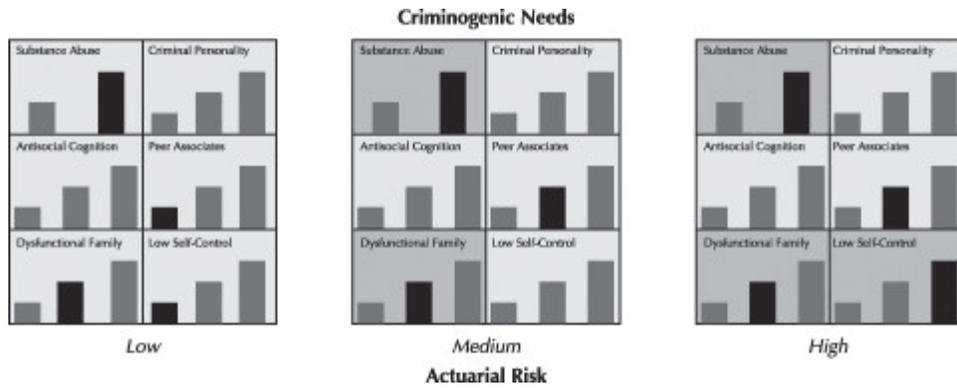
Criteria for Antisocial personality Disorder

Antisocial personality disorder is defined as a pervasive pattern of disregard for and violation of the rights of others since the age of 15 as indicated by three or more of the following:

1. Failure to conform to social norms with respect to lawful behaviors 2. Deceitfulness (repeated lying) or use of aliases or conning others for personal profit or pleasure
2. Impulsivity or failure to plan ahead
3. reckless disregard for safety of self or others
4. Consistent irresponsibility, as indicated by failure to sustain work and honor financial obligations
5. Lack of remorse, as indicated by being indifferent to or rationalizing having hurt, mistreated, or stolen from another

Source: ApA, 2004

Exhibit 1: Individual Criminogenic Needs & Actuarial Risk: Identifying the Pattern to Determine the Intervention



[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 MMASOR) 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

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

[Home](#)

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

[References](#) | [Endnotes](#)

Stephen D. Gottfredson

Laura J. Moriarty

Virginia Commonwealth University

[Probation Officers and Correctional Treatment Specialists](#)

[Human Judgment](#)

[The Place for Personal Judgment](#)

[Conclusion](#)

IN VIRTUALLY ALL decision-making situations that have been studied, actuarially developed devices outperform human judgments. This is true with respect to psychiatric judgments (see, for example, Meehl, 1965; Gough, 1962; Ennis and Litwack, 1974); graduate school admissions (e.g., Dawes, 1979; Dawes and Corrigan, 1974); prognostic judgments made by sociologists and psychiatrists relative to a parole-violation criterion (Glaser, 1955, 1962); parole board decisions (Gottfredson, 1961; Gottfredson and Beverly, 1962; Carroll, Wiener, Coates, Galegher, & Alibrio, 1982); mental health and correctional case worker judgments of offender risk (Holland, Holt, Levi, & Beckett, 1983), spousal assault (Hilton and Harris, 2005) and in other areas (Goldberg, 1970), including the analysis of credit risk (Somerville and Taffler, 1995). Indeed, a recent review and meta-analysis of 56 years' accumulation of research on the "clinical vs. statistical" prediction "problem" conducted as part of a Festschrift for Paul E. Meehl, a pioneer in the field, again confirms that statistical models outperform clinical decision-makers (Ægisdóttier, White, Spengler, Maugherman, Anderson, Cook, Nichols, Lampropoulos, Walker, Cohen and Rush, 2006).

The relative superiority of statistical to intuitive methods of prediction is due to many factors. For example, human decision-makers often do not use information reliably (e.g., Ennis and Litwack, 1974), they often do not attend to base rates (Meehl and Rosen, 1955), and this has been specifically illustrated in criminal justice decision-making (Carroll, 1977); they may inappropriately weight items of information that are predictive, or they may assign weight to items that in fact are not predictive; and they may be overly influenced by causal attributions (e.g., Carroll, 1978) or spurious correlations (Monahan, 1981). In fairness, it should be pointed out that there may be advantages to intuitive judgments as well. For example, human decision-makers can make use of information that cannot be made available to a statistical device (at least readily). Demeanor during an interview may be one such example. Other factors in favor of intuitive judgments are reviewed in Dawes (1975; Dawes, Faust, and Meehl, 1989). [□](#)

Given these facts, is there reason to still consider clinical judgments when determining risk-assessment within a justice system population? Indeed, with the 1998 publication of *Violent Offenders: Appraising and Managing Risk* (Quinsey, Harris, Rice and Cormier), we find an argument that we should not. "What we are advising is not the addition of actuarial methods to

existing practice, but rather the complete replacement of existing practice with actuarial methods” (p. 171; see Litwack, 2001 for a strong rebuttal in the arena of the assessment of dangerousness). We argue that even though statistical prediction is superior to clinical judgment in almost all settings, this does not obviate the need for nor value of clinical judgment in a variety of arenas, including some criminal justice venues. We use the roles of probation officers and correctional treatment specialists to provide examples.

[back to top](#)

Probation Officers and Correctional Treatment Specialists

Among the largest group of criminal justice professionals is that working in corrections. And with the vast number of adults and juveniles on probation, parole or incarcerated, the workload of these individuals is quite high.

According to the U.S. Department of Labor, Bureau of Labor Statistics, there are about 90,600 probation officers and correctional treatment specialists nationally (Bureau of Labor Statistics, 2006), and in the federal system, there are approximately 5,000 officers throughout the United States and its territories (personal communication, Richard Gayler, May 31, 2006). The number of adults on probation in 2004 was about 4.1 million (Glaze and Palla, 2005), for an average caseload nationally of about 46. The top three states in terms of employment of probation officers and correctional treatment specialists are California (13,090), Texas (6,100) and Florida (5,760). When examining data from 2004 that reports the state’s community corrections population, we find that Texas has the largest, with 428,773 adults under supervision, followed by California with 384,852, and Florida with 281,170 (Glaze and Palla, 2005). Using these figures, the average caseloads range from a high of 70 to a low of 29.

Qualifications for employment as a probation officer or correctional treatment specialist vary by state, but a Bachelor’s degree in social work, criminal justice or some other related field typically is required (Bureau of Labor Statistics, 2006). Some states require a more advanced degree—Master of Science or Master of Arts in a related field (psychology, sociology, criminology, etc.), often with an additional experiential requirement as well. Many states require that probation officers and correctional treatment specialists receive training, upon completion of which the candidate must pass a certification test. Typically, new officers work as trainees or for a probationary period before they become permanent employees.

Probation officers supervise offenders who have been placed on probation while correctional treatment specialists counsel and create rehabilitation plans for offenders to follow when they are no longer incarcerated or on parole (Bureau of Labor Statistics, 2006). Probation officers also spend a great deal of time in court, investigate offender backgrounds, write pre-sentence investigation reports, and recommend sentences and treatment plans. Correctional treatment specialists may work in jails, prisons, or probation or parole agencies, where they might evaluate the progress of inmates, develop parole and release plans, write case reports for parole boards and other decision-makers, and/or develop and write treatment plans and summaries for clients.

What we have then, is a large number of highly qualified and trained professionals who routinely are required to make prognostic decisions about offenders. Elsewhere, we have described ways to improve the reliability of these decision-making processes (Gottfredson and Moriarty, 2006). We have argued that the use of actuarial devices invariably increases the reliability and prognostic validity of decisions made in these settings (Gottfredson and Moriarty, 2006), and, as noted above, some would argue that the surest way to do this is to rely solely on statistical prediction, such as risk-assessment tools, as a way to increase the accuracy and reliability of these decisions (Quinsey, Harris, Rice and Cromier, 1999). Is it indeed time to supplant human judgment in justice system settings with the cold calculus of the actuary?

[back to top](#)

Human Judgment

Judgments are made routinely in a host of fields including psychiatry and psychology (Kleinmuntz, Faust, Meehl, & Dawes, 1990; Dawes et al., 1989); mental health (Ægisdóttir et al., 2006); dangerousness (Litwack, 2001); economics (Dawes, 1999); forecasting (Bunn & Wright, 1991), medicine, engineering, finance, management (Kleinmuntz et al., 1990, p. 146); interpersonal violence (Hilton, Harris & Rice, 2006; Mills, 2005); and forensics (Hilton, Harris, Rawson & Beach, 2005; Harris, Rice & Cormier, 2002), in addition to those noted earlier in this paper. In most cases, the literature reveals strong support for the accuracy of actuarial prediction over human judgment. This is a longstanding finding, replicated in dozens of venues (Dawes et al., 1989; Kleinmuntz et al., 1990; Westen & Weinberger, 2005). As noted earlier, there are many reasons to expect that actuarial methods will outperform human judgments. In addition to those reasons cited above, these methods may be expected to provide other benefits:

Even when actuarial methods merely equal the accuracy of clinical methods, they may save considerable time and expense. ... When actuarial methods are not used as the sole basis for decisions, they can still serve to screen out candidates or options that would never be chosen after more prolonged consideration. When actuarial methods prove more accurate than clinical judgment the benefits to individuals and society are apparent (Dawes et al., 1989, page 1673).

Why, then, should we continue to allow (indeed, require) probation officers and case managers to exercise individual discretion, when an actuarially-derived tool may be expected to perform better? As Westen and Weinberger (2005) remind us in a discussion of the pioneering work of Paul E. Meehl on clinical and statistical prediction, even though statistical prediction will routinely outperform clinical prediction, we should not lose sight of the fact that “actuarial procedures are far from infallible, sometimes achieving only modest results” (Dawes et al., 1989, p. 1673). [For a discussion of the methodological and statistical problems associated with such applications and the resultant fallibilities of such procedures, see Gottfredson and Moriarty, 2006]. Still, as Dawes (2005) concludes, “whenever statistical prediction rules (SPR’s) are available for making a relevant prediction, they should be used in preference to intuition” (p. 1245).

But does the superiority of actuarial procedures over clinical judgment mean that there is no place for clinical judgment in predicting behavior? The answer is no: “an enormous amount of prediction is made by human judgment” (Darlington, 1986, p. 362). Simply put, clinical methods of decision-making rest in the decision-maker’s head, while statistical or actuarial methods eliminate the human judgment with the “conclusions rest(ing) solely on empirically established relations between the data and the condition or event of interest” (Dawes et al., 1989, p. 1668). There are instances when clinicians can make valid inferences (Westen and Weinberger, 2005), and there are times when it is preferable to use both clinical and statistical judgments to predict behavior. As Darlington (1986, 362) reports, “This research does not suggest human judgment is generally unnecessary; rather it indicates that the most accurate predictions generally result from a predictive system in which human judgment and statistical analysis are mixed according to prescribed rules.” Moreover, Dawes et al. (1989) assert that there are instances when clinical judgment might improve the actuarial method. They cite specifically the following examples: “judgments mediated by theories and hence difficult or impossible to duplicate by statistical frequency alone; select reversal of actuarial conclusions based on the consideration of rare events or utility functions that are not incorporated into statistical methods; and complex configurational relations between predictive variables and outcome” (p. 1670).

[back to top](#)

The Place for Personal Judgment

As some assert, the dilemma once posed by “using either the head or the formula is no longer the main focus of contemporary decision research. Rather, the focus has long ago shifted to evaluating the use of both modes of information combination in tandem” (Kleinmuntz et al., 1990, p. 146; and see generally, Litwack, 2001).

In some of the most recent research examining violence risk assessment (Hanson, 2005), we find that the question has shifted from whether violence can be predicted, to what is the best method of risk assessment. The validation research typically found that a series of measures reviewed in the article showed moderate accuracy in predicting violent recidivism. The question then is: Might the prediction have been improved if clinical judgments were included as well?

This question is partly answered by Douglas, Yeoman, and Boer (2005), who studied violence risk in a sample of criminal offenders. Douglas and colleagues looked at the predictive validity of multiple indices of violence risk. Although they conclude that “several indices were related to violent recidivism with large statistical effect sizes,....” they also found the findings to be “inconsistent with a position of strict actuarial superiority, as HCR-20 [2](#) structured risk judgments—an index of structured professional or clinical judgments—were as strongly related to violence” (479).

[back to top](#)

Conclusion

With very few established exceptions, statistical prediction clearly outperforms clinical judgment. Accordingly, we certainly would not advocate use of clinical judgment over statistical prediction. For as Dawes reports,

The superiority of statistical prediction is crystal clear when clinical judgment is pitted against actuarial analysis in a situation where both are based on the same information—so that the problem is basically one of how to combine it. It also has been found that clinical judgment in psychology is inferior in situations where the important variables captured by the statistical model constitute a proper subset of the variables considered by the clinician. It is also true that the statistical models need not even be optimal. Nevertheless, clinical psychologists make a great deal of money by relying on their intuitions for combining information and for making predictions, and in courts they eschew statistical models, instead proudly proclaiming that “in my experience...” What happens here is that the “inside view” is preferred to the outside one, despite massive evidence that that outside one is superior (Dawes, 1999, pp. 37-38).

However, there are times when a combination of the two may better serve clientele. As Dawes et al. (1989) report, “Clinicians might be able to gain an advantage by recognizing rare events that are not included in the actuarial formula (due to their infrequency) and that countervail the actuarial conclusion (p. 1670).

And while such incidents might be infrequent, it is also true that the probation officers and correctional counseling specialists must have a role in decision-making that goes beyond the mere administering of the risk-assessment devices. There is a place for human judgment and experience in the decision-making process, and we must value their continued consideration.

However, as noted by Sir Frances Bacon, “We do ill to exalt the powers of the human mind, when we should seek out its proper helps” (as quoted in Hogarth (1980)). In light of the well-known tendency for justice system decision-makers to concentrate on information that is demonstrably not predictive of offender behavioral outcomes (Gottfredson and Gottfredson, 1986), and the potential consequences of this for affecting the validity of prognostication (Gottfredson and Moriarty, 2006), caution is the order of the day.

[back to top](#)

[References](#) | [Endnotes](#)

not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

[Home](#)

Overcoming Sisyphus: Effective Prediction of Mental Health Disorders and Recidivism Among Delinquents

[References](#)

Albert R. Roberts, Ph.D.

Professor of Criminal Justice, Rutgers University

Kimberly Bender, M.S.W.

Doctoral Candidate, University of Texas at Austin

[Juvenile Assessment Centers](#)

[Components of Risk and Need Assessment](#)

[Tools for Mental Health and Associated Risk Assessment](#)

[Conclusion and Suggestions for Further Research](#)

IN THIS AGE OF accountability and performance-based measures, criminal justice professionals are being increasingly required by state and federal agencies to demonstrate the reliability and validity of their assessment instruments including brief symptom inventories, diagnostic tools, and violence risk assessment measures. Risk assessment tools assist institutional classification boards as well as parole boards to: 1) determine an initial security rating and placement into a particular facility and program(s); 2) develop a rehabilitation treatment plan; 3) assess eligibility for early release; and 4) determine the type of supervision needed while on parole. This article first describes how the juvenile justice system assesses youths' risks and needs through Juvenile Assessment Centers, then explores common components of assessment in the juvenile justice system, and concludes with an examination of the most commonly used risk and mental health assessment tools and the evidence that supports their use.

All experienced probation officers, juvenile counselors, and forensic clinicians should have skills in risk assessment. Clinical assessment knowledge and skills provide the foundation for clinical judgments, applied research, and evidence-based practice. Within the juvenile justice system, prediction can be operationally defined as an assessment of future lawbreaking for the juvenile offenders who are officially processed through the system.

There are two primary types of prediction: clinical and actuarial. Clinical predictions are made by trained juvenile justice and forensic specialists after they have examined an individual's criminal and psychosocial history, and the results from psychosocial scales and inventories. Actuarial prediction methods are based on known properties, parameters and statistical formulas applied to identical sets of data (e.g. demographic data, criminal history). Because of the two authors' backgrounds in forensic mental health and social work, we focus on clinical judgments and the most commonly used assessment scales for measuring mental health status, psychosocial functioning, and future criminality.

There is no single scale or assessment tool that can predict future mental health status or criminality with 100 percent certainty. Behavior, abilities, peer influences, family factors and

deviant behavior patterns are not static. They often change with age and different experiences. Empirical evidence from classic longitudinal studies indicate that violent juveniles are strongly influenced by male siblings of similar ages, delinquent gangs, and small groups of delinquent friends (Farrington & Loeber, 2000; Farrington & West, 1990). Therefore, it is critically important to use multiple assessment tools with clients at different points in the juvenile justice process.

Clinical prediction is based on perceptions and judgments in which the juvenile justice professional and/or mental health clinician uses different data sources, such as clinical diagnosis, ratings and scores on psychosocial risk assessment scales, interviews, psychotherapy records, and criminal history data to make judgments about the offender's placement in institutional or community-based treatment programs, progress, and discharge from probation.

The early roots of prediction in juvenile justice can be traced to the establishment of the first juvenile court mental health clinic in Chicago in 1909 (Roberts, 2004), and the rapid growth and development of over 600 child guidance clinics by the late 1950s connected to juvenile courts throughout the United States (Roberts, 2004). In these clinics forensic psychiatrists and social workers collaborated on behalf of troubled juveniles.

By the late 1990s, most state juvenile correctional agencies included formal and informal dangerousness and risk of further violence and re-offending in their intake classification and assessment centers. The goal of risk assessments is twofold: 1) to predict the probability that a juvenile offender will re-offend; 2) to predict which youths are at high risk of exhibiting violence in the institution or residential treatment facility, or upon release to parole supervision in the community. In general, classification decisions are made based on forecasts regarding which treatment/rehabilitation program is likely to be effective in changing the behavior patterns of specific types of juveniles, generally viewed as either property-related offenders or violent offenders adjudicated for offenses against persons.

One of the most overlooked areas of juvenile justice is the assessment and treatment of juvenile offenders with mental health disorders, especially co-morbid psychiatric disorders. Research indicates that at least two-thirds of juvenile detainees have one or more mental health disorders in addition to their juvenile offenses. Incarcerated juveniles suffering from impulsiveness, hopelessness and depression are at an increased risk of suicide ideation, suicide attempts, and death (Rapp-Pallicchi & Roberts, 2004). The death rate from suicide is 4.6 times higher in juvenile detention centers than in the general population (Sheras, 2000). Therefore, it is imperative that experienced mental health professionals be hired by juvenile justice agencies so that they can conduct extensive assessments at the pre-adjudicatory, incarceration, and community release stages (Rapp-Pallicchi & Roberts, 2004). At the present time, most large juvenile probation departments do have a few probation officers with expertise in forensic mental health assessment and treatment. However, the time is now ripe for the National Juvenile Detention Association (NJDA), as well as the American Correctional Association (ACA), and state and county correctional administrators to follow the lead of the American Probation and Parole Association (APPA) in giving priority to setting standards, and encouraging their members to hire and train staff in all aspects of juvenile assessment and treatment.

[back to top](#)

Juvenile Assessment Centers

A promising advancement of juvenile assessment is the innovative development of centralized, single point of entry, intake juvenile assessment centers. These assessment centers are based on a general model for bringing together a variety of community agencies to one centralized location in which all justice system-involved youth can receive thorough assessment. Juvenile justice, law enforcement, school truancy, diversion programs, and other human service agencies are centrally located, allowing for efficient and comprehensive assessment of youths' risks and service needs (Dembo, Schmeidler, & Walters (2004).

Key elements of juvenile assessment centers (JACs) include 1) A single, 24-hour, centralized point of contact for all youth in contact or at risk of contact with the juvenile justice system, 2) Screenings and comprehensive assessments of youths' circumstances and service needs, 3) Management Information Systems that centralize information to avoid repetition and assure appropriate treatment, and 4) Case management services that integrate information in order to recommend appropriate referrals and follow up on youth after they are referred (Dembo, Schmeidler, & Walters, 2004).

Juvenile assessment centers got their start in the early 1990s in Florida and quickly gained the attention of the Florida legislature, which was struggling with prison overcrowding. With a growing budget due to special appropriations, JACs quickly spread to several counties across Florida and were eventually established in other states, including Colorado and Kansas. Investing further in assessment centers with an initiative in 1996, the OJJDP allocated funds to two assessment centers, in Denver, CO and Lee County, FL, designated as planning sites to develop more assessment centers. Additional funds supported improving services at two designated enhancement sites in Jefferson County, CO and Orlando, FL (Dembo, Schmeidler, & Walters, 2004).

JACs vary considerably by location, due in large part to access resources and the unique needs of the communities they are serving. For example, many Florida JACs work closely with nearby juvenile addiction receiving facilities to provide detoxification, assessment, and stabilization for youth with substance abuse problems. JACs differ in the range of services they provide, from those with only juvenile justice agencies to those such as the Hillsborough County, Florida JAC that provides an array of services, including booking, supervision, detention center screening, diversion, and truancy programming at one site. JACS located in urban settings tend to have longer hours, process many youth, and thus conduct more thorough assessments off-site (Dembo, Schmeidler, & Walters, 2004).

Despite these differences, JACs share common benefits to the juvenile justice system. They provide a centralized site for legally required mandates to be carried out more efficiently, saving time locating youth, completing multiple screenings, and providing information to courts for decision making. Integrating information in one information system allows for better-informed decisions regarding need for services and necessary level of supervision. Accessing all system-involved youth, JACs create a prime opportunity for prevention and early intervention. Finally, on a macro-level, information from JACs informs the community of broader juvenile justice trends and needs for new services (Dembo, Schmeidler, & Walters, 2004).

Dembo et al. (2004) note an ongoing struggle for funding experienced by many JACs. Consistent funding at the federal and state levels is needed in order to provide decent salaries to well-trained staff, thereby reducing staff turn-over and improving quality of service. Additional funds would also allow JACs to maintain their original goals of comprehensively responding to youths' multifaceted needs, preventing JACs from skimming services and becoming mere processing centers.

With necessary funding and support, the future utility of JACs is broad and influential. JACs have the potential to play a major role in developing empirical knowledge in the future. With large sample sizes, JACs' information systems could easily gather data on youths' characteristics, service needs, and outcomes in different treatment programs, providing juvenile justice research with difficult to obtain information. This information can then be used to inform program development and service provisions to juvenile offenders.

JACs can also provide much-needed solutions to assessment, referral, and service delivery in the future. By integrating information among many agencies, JACs can help to identify youth who slip through the system by failing to follow through on treatment recommendations. Furthermore, providing objective measures of substance use through urinalysis screening is an invaluable service offered through JACs and has implications for validating youths' self reports of substance use and subsequent appropriate treatment placements. Finally, JACs ensure investment in prevention efforts, keeping youth from further developing delinquency careers; these prevention

efforts inversely relate to the number of youth requiring long-term incarceration that is expensive and fairly ineffective (Dembo, Schmeidler, & Walters, 2004).

[back to top](#)

Components of Risk and Need Assessment

The central goals of youth assessment in juvenile justice are: 1) the safety of the community by preventing re-offending; and 2) youth rehabilitation and clinical treatment. In other words, mental health assessments seek to identify both *risk and treatment needs*. Assessments must be comprehensive and cover several domains. Comprehensiveness includes assessing a youth's offense history, family/environmental factors, education/employment history, peer relationships, and psychosocial functioning.

Assessing psychosocial functioning is particularly important as the juvenile offender population has elevated rates of mental health and substance use disorders (Teplin, Abram, McClelland, Duclan, & Mericle, 2003). Furthermore, these psychosocial factors (i.e. personality characteristics, behaviors, affect, attitudes, beliefs and interpersonal constructs) predict youths' infractions while incarcerated and their behaviors once they are released into the community (Cauffman, 2004; Hathaway & Moncachesi, 2003).

Mental health status is often under-assessed and consequently under-treated in the juvenile justice system. This is because of a lack of resources and trained staff, as well as a punishment mentality. Teplin et al. (2005) report that only 15.4 percent of detained adolescents who needed mental health treatment received treatment in the detention center; it is estimated that as many as 13,000 detained youth with major mental health disorders go untreated (Teplin et al., 2005). Effective mental health assessment and treatment are critical for achieving effective juvenile justice.

In 2002 the Consensus Conference convened, composed of more than 20 researchers with expertise in mental health assessment and juvenile justice, with the aim of developing recommendations for mental health assessment in the juvenile justice system. The Consensus Conference brought together nationally recognized experts in the areas of mental health, juvenile justice, and child welfare service systems. It was guided by data from a national survey of current mental health assessment practices conducted by the Center for the Promotion of Mental Health Assessment in Juvenile Justice. Directed by Gail Wasserman at Columbia University's Department of Child Psychiatry, the Center's national survey provided information on the current practices and needs of juvenile justice systems across the nation. From these findings, the Consensus Conference was then able to create recommendations for standardizing mental health assessment practices on a national level. The Consensus Conference recommended that these four types of assessments should be conducted:

1) Emergent risk needs should be assessed immediately upon arrival at a secure facility; 2) A comprehensive mental health assessment should be conducted on all youths at the facility to identify those needing more thorough mental health assessments; 3) Prior to community re-entry, all youth should be assessed to facilitate transition and referral to community mental health services; and 4) continued re-assessments should take place after the youths have returned to the community, to prevent re-offenses.

In the past two decades, several measures have been developed to assess juvenile offenders' mental health and associated risks (Grisso, 2005). These measurement instruments aim to be accurate, reliable, and thorough while being fairly quick and inexpensive to administer.

[back to top](#)

Tools for Mental Health and Associated Risk Assessment

There are currently several well-validated assessment measures used to predict the likelihood of re-offending upon release, mental health treatment needs, and danger towards self (suicide

ideation and suicide attempts) and others, based on the presence or absence of substance abuse, suicide ideation, personality traits, thought disturbance, and depression-anxiety. Below we describe several of the most common assessment tools used in juvenile justice research and practice. Several scales are actuarial in nature while others integrate actuarial assessment with supplemental clinical judgment. Instruments are categorized according to their utility as brief screening tools, comprehensive assessment instruments, or risk assessments predicting recidivism or dangerousness in the future. Descriptions are intended to give a brief overview and should not be considered full reviews. For more detailed information on each of these instruments, readers are directed to Grisso, Vincent, and Seagrave's (2005) *Mental Health Screening and Assessment in Juvenile Justice* or to literature by each scale's developer.

Brief Screening Tools

Brief screening tools are instruments that can be administered very quickly (usually in 30 minutes or less) and help staff to identify youth who may be of immediate risk to self or others. Furthermore, the screenings should help staff identify youth in need of more comprehensive mental health assessment. These instruments should be easily administered by front-line staff with little specialized training, allowing for quick and inexpensive use. Brief screening tools should not be used to inform treatment plans; instead their utility is in identifying those youth in need of emergency mental health services or those who need more comprehensive assessment that can then inform treatment needs. [Table 1](#) describes the strengths and limitations of three commonly used brief screening tools.

MAYSI-2. The Massachusetts Youth Screening Instrument—Version 2 (MAYSI-2) was developed by Grisso and Barnum (2003) as a self-report measure to identify youth entering the juvenile justice system with thoughts, feels, or behaviors indicative of mental health problems. The MAYSI-2 can be administered by pencil-paper or by CD-ROM and consists of 52 yes-no questions asking whether each item is true for the youth. Seven subscales are assessed, including alcohol/drug use, angry-irritable, depressed-anxious, somatic complaints, suicide ideation, thought disturbance, and traumatic experiences. This objective measure includes cut-off scores from a normative juvenile justice sample that can be used as indicators of clinical significance (Grisso & Quinlan, 2005). Research evaluating the reliability of the MAYSI-2 reports internal consistency ranging from .61 to .86 (Grisso et al., 2001) and support for test-retest reliability on most subscales (Cauffman, 2004). Similar positive findings were found in studies of validity comparing the MAYSI-2 to other standardized scales (Espelage et al., 2003) and to the DSM-IV (Wasserman et al., 2004). Of interest were several studies that found the MAYSI-2 to predict future behaviors such as institutional maladjustment, sentence length, and necessary intervention for suicide risk and assaultive behavior (Cauffman, 2004; Stewart & Trupin, 2003). Cauffman and MacIntosh (2006) recently found different properties on some subscales, in particular the alcohol/drug use, anger-irritability and suicide ideation subscales, across ethnic and gender groups. Further research should continue to examine the extent to which these subscales are valid measures for female and ethnic minority youth.

POSIT. The Problem-Oriented Screening Instrument for Teenagers (POSIT) was developed by Rahdert (1991) as a self-report brief screening to identify troubled youths' problems in psychosocial functioning requiring further assessment. The POSIT, available by pencil-paper or by CD-ROM, consists of a self-administered questionnaire with 139 yes-no questions and assesses 10 functional areas, including substance use/abuse, physical health, mental health, family relations, peer relations, educational status, vocational status, social skills, leisure/recreation, and aggressive behavior/delinquency (Dembo & Anderson, 2005). Youths' total scores in each problem area can be compared to empirically-based cut-off scores allowing for a classification of low-, medium-, or high-risk for that problem area. While the POSIT is objectively scored, collateral information is recommended to validate youths' responses. Research evaluating the reliability of the POSIT indicates internal consistency exceeded .70 and test-retest reliability significantly better than chance (Knight, Goodman, Pulerwitz, & DuRant, 2001). Hall, Richardson, Spears, & Rembert (1998) found high construct validity for the POSIT. Preliminary research indicate the POSIT is useful in classifying youth by predicting return to the juvenile justice system (Dembo, Turner, et al., 1996).

CAFAS. The Child and Adolescent Functional Assessment Scale (CAFAS) was developed by Hodges (2000a) to assess youths' everyday psychosocial functioning across school, home, community, and work settings. Different ratings (parents, teachers, youth) of youth's behaviors are obtained across 10 subscales (school/work, home, community, behavior toward others, moods/emotions, self-harmful behavior, substance use, thinking, material needs, and family/social). Questions are asked for each subscale that identify severe, moderate, mild, or no impairment and also include questions that indicate strengths or protective behaviors exhibited by the youth. Raters score the CAFAS after collecting information based on both their own observations and a family of instruments that assess the youth's and their caregiver's perspective on everyday functioning. Studies report positive results for test-retest reliability, validity, and ability to predict level of service needs (Hodges & Wong, 1997).

Comprehensive Assessment Instruments

Before forensic mental health specialists, correctional counselors, and probation officers can recommend a treatment plan, comprehensive risk assessment data must be collected. In comparison to brief assessment tools, comprehensive assessment instruments more thoroughly assess several domains of youths' mental health, personality, and psychosocial characteristics. These assessments often involve longer, more intensive interviews and several also collect collateral information from other settings in the youth's life (i.e., teachers, parents, or chart information). Comprehensive assessment instruments help to clarify mental health needs, can inform treatment planning, and are most often conducted by professionals or require more involved training. [Table 2](#) describes the strengths and limitations of three commonly used comprehensive assessment instruments.

DISC. The Diagnostic Interview Schedule for Children: Present State Voice Version (Voice DISC) was developed by Shafer, Fisher, Lucas, Dulcan, & Swab-Stone (2000) to assess mental health problems and provides a diagnosis by evaluating how youth meet DSM-IV criteria. Now self-administered on the computer, the assessment employs a unique pattern of questions based on respondents' answers to previous questions, assessing the degree to which they meet criteria for more than 30 diagnoses (Wasserman, McReynolds, Fisher, & Lucas, 2005). Subscales include anxiety, mood, disruptive behavior, substance use, and miscellaneous (eating disorders, tic disorders, etc.). After the assessment tool determines a youth meets diagnostic criteria, further questions inquire about the severity and frequency of these problems in an attempt to understand impairment. However, youth may be limited in their ability to recognize the consequences of their own behaviors, and it is suggested that clinicians use collateral information to determine impairment. DISC reports include a list of those diagnoses for which the youth met criteria, impairment and symptom scores, and a list of "clinically significant symptoms." Acceptable reliability for most diagnoses and good test-retest reliability have been reported (Shaffer et al. 2000). Moderate to poor correlation with clinician diagnosis has been found (Aronen, Noam, & Weinstein, 1993); however, independent clinical diagnosis is known to be fairly subjective and unreliable.

MMPI-A. The Minnesota Multi-phasic Personality Inventory—Adolescent (MMPI) was developed by Archer (1997) and adapted for adolescents by Butcher et al. (1992) and is the most widely used personality assessment (Archer & Baker; 2005). The MMPI consists of 478 items with validity scales (e.g. defensiveness, tendency to exaggerate, response consistency), clinical scales (e.g. psychopathology such as depression, anxiety, schizophrenia, antisocial behaviors), content scales (e.g. externalizing behaviors, anger, low self-esteem), and supplementary scales (immaturity, repression). Raw scores are converted to t-scores and are compared to normative scores, resulting in classification of youth who are clinically elevated, marginally elevated, or typically adolescent. Early research found scale 4 (psychopathic deviate) especially helpful in predicting delinquency (Hathaway & Monachesi, 1963). This is confirmed in later studies that found scale 4 (psychopathic deviate), scale 8 (schizophrenia) and scale 9 (Hypomania) predictive of higher rates of delinquency (Archer, Bolinsky, Morton, & Farris, 2003). Over 100 studies have examined aspects of the MMPI-A, and the instrument is known for its good reliability and validity ; a good resource is a review by Forbey (2003).

MACI. The Millon Adolescent Clinical Inventory (MACI) developed by Millon (1993) as a short assessment that provides clinical information on a variety of psychological problems, including psychopathology, peer difficulties, family problems, and confusion about self (Salekin, Leistico, Schrum, Mullins, 2005). It also assesses a balance of externalizing/delinquency risk factors as well as suicidal tendency and risk towards self. Based on the DSM-IV, the MACI includes: 3 validity scales (disclosure, desirability, debasement), a reliability scale, 7 clinical syndrome scales (eating dysfunction, substance abuse, delinquent predisposition, impulsive propensity, anxious feelings, depressive affect, suicidal tendency), 12 personality scales (introversive, inhibited, doleful, submissive, dramatizing, egotistic, unruly, forceful, conforming, oppositional, self-demeaning, borderline tendencies), and 8 expressed concern scales (identity confusion, self-devaluation, body disapproval, sexual discomfort, peer insecurity, social intensity, family discord, child abuse). Base-rate scores are calculated and are interpreted by mental health professionals, who first examine validity and reliability before identifying problem scales with elevated base-rate scores. The MACI is shown to have good internal and test-retest reliability and concurrent and predictive validity (Millon, 1993). A recent study by Taylor, Skubic-Kemper, Loney, and Kistner (2006) extends support for using the MACI as a tool for classifying subtypes of serious juvenile offenders. Furthermore the MACI has been shown useful in assessing clinical change from intake to discharge in inpatient settings (Piersma, Pantle, Smith, Boes, & Kubiak, 1993) and is predictive of recidivism (Salekin, Ziegler, Larrea, Anthon, and Bennet (2003)).

Risk for Recidivism and Dangerousness Assessment Tools

Research has identified several factors that put youth at risk for future violence or recidivism. While no definitive list of factors has been developed, research has shown that there are common pathways to recidivism that can be predicted with some accuracy. These factors have been used to compose assessment tools that measure youths' risk of re-offending once released into the community. These instruments help juvenile justice centers make decisions to protect the community and identify need for further services such as case management. Risk for recidivism assessments often involve collecting collateral information from parents or chart materials in addition to interviewing, and is time-intensive. Thus, these assessments require specialized training or a professional degree to administer and score. Assessments of future risk behavior include varying degrees of clinical judgment to interpret the results and make decisions. [Table 3](#) describes the strengths and limitations of three commonly used assessments of future risk.

YLS/CMI. The Youth Level of Service/Case Management Inventory (YLS/ CMI; Hoge, Andrews, & Leschied, 2002) is designed to predict juvenile offender recidivism as well as case management needs, making it especially useful in planning for transitions out of the juvenile justice system. The YLS/CMI assesses the offender as high or low risk, assesses need by targeting services due to risk factors, and assesses responsivity or reaction to interventions. The YLS/CMI is composed of six sections: 1. Assessment of risk and needs (42-item checklist assessing prior/current offenses, family circumstances, education/employment, peer associations, substance abuse, leisure, personality/behavior, and attitude); 2. Summary of risk/need factors (comparing scales to normative ranges); 3. Other needs/special circumstances (situational information such as parental drug use or behavioral records that add information specific to youth); 4. Professional override feature (asks clinician to use clinical judgment considering all relevant information to rate youth's risk level); 5. Contact level (intensive services should be recommended for high risk youth); and 6. Case management plan (specific goals and objectives for reaching goals). Due to the complexity and knowledge it requires, the YLS/CMI is completed by a trained professional and purposely incorporates a degree of clinical judgment to supplement the objective portions of the assessment. Adequate internal consistency (Rowe, 2002) and inter-rater reliability have been found in empirical studies (Schmid, Hoge, & Robertson, 2002), except for the leisure/recreation subscale, which has a wide range of interrater reliability (.05-.92). Several subscales of the YLS/CMI have been correlated with other externalizing measures (Rowe, 2002). Ability to predict new charges, new convictions and serious offense charges have been consistently demonstrated with males and more inconsistently for girls (Rowe, 2002; Schmidt et al., 2002).

SAVRY . The Structured Assessment of Violence Risk in Youth (SAVRY; Bartel, Borum, & Forth, 2000) involves professional judgment based on systematic appraisal of the degree to which youth demonstrate risk factors for future violence. The appraisal involves assessment of 6 protective factors (prosocial involvement, strong social support, strong attachments and bonds, positive attitude toward intervention and authority, strong commitment to school, resilient personality traits). The instrument also assesses 24 risk factors including: historical (history of violence, of nonviolent offending, early initiation of violence, history of self harm, childhood exposure to maltreatment, parental criminality, early caregiver disruption, poor school achievement), individual (negative attitudes, risk taking/impulsivity, substance use difficulties, anger management problems, low empathy/remorse, ADHD difficulties, poor compliance, low interest/commitment to school) and social/environmental (peer delinquency, peer rejection, stress/poor coping, poor parental management, lack of personal/social support, community disorganization) domains. Information should be gathered by the examiner through interviews with the youth, review of records, and observation (Borum, Bartel, Forth, 2005). Numerical ratings are not the goal of this assessment; identifying empirically validated risk factors specific to each youth is the goal. Thus clinicians are faced with reviewing the identified risk factors and making a clinical judgment about a youth's overall risk. Inter-rater reliability is moderate to high (.81) (Catchpole & Gretton, 2003) and studies show support for concurrent validity as compared to the YLS/CMI and PCL:YV (Catchpole & Gretton, 2003). Moderate yet significant correlations were found between the SAVRY and measures of violence and aggression (McEachran, 2001; Gretton & Abramowitz (2002). Additionally, those youth characterized as low risk had violent recidivism rates (6 percent) much lower than those characterized as moderate (14 percent) or high risk (40 percent) (Catchpole & Gretton, 2003).

PCL:YV . The Hare Psychopathy Checklist: Youth Version (PCL:YV; Forth, Kosson, & Hare, 1990) uses multiple sources of information across interpersonal, affective and behavior domains to identify symptoms predictive of serious psychopathy in adolescents. The examiner uses information from an intensive interview with the youth, collateral sources, and review of the chart to rate the youth according to a 20-item checklist including: impression management, grandiose sense of worth, stimulation seeking, pathological lying, manipulation of personal gain, lack of remorse, shallow affect, lack of empathy, parasitic orientation, poor anger control, impersonal sexual behavior, early behavior problems, lacks goals, impulsivity, irresponsibility, failure to accept responsibility, unstable interpersonal relationships, serious criminal behavior, serious violence of conditional release, and criminal versatility. Total scores provide the number of psychopathic features observed for each youth but do not result in cut off or classification. However raters can compare youth's scores to percentile scores based on institutional, probation, and community samples. After extensive training required to administer the PCL:YV, inter-rater reliability scores are generally high (.90-.96) and internal consistency is adequate (.85-.94, Forth et al., 2003). Moderate correlations with reports of delinquency, externalizing symptoms and aggression are reported, while the PCL: YV (as intended) does not correlate with measures of internalizing disorders (Cambell et al., 2004). Recent studies report the PCL:YV significantly predicted both violent and nonviolent recidivism (Corrado, Vincent, Hart, & Cohen, 2004) as well as clean urine screens and participation in treatment (O'Neill et al., 2003a). Eden, Buffington, Colwell, Johnson, & Johnson (2002) further support the ability of the PCL: YV to predict disciplinary infractions in their sample juvenile sex offenders. However, Spain, Douglas, Poythress, & Epstein (2004) found negative results, with no relationship evident between the PCL:YV and treatment progress.

[back to top](#)

Conclusion and Suggestions for Further Research

Several Brief Assessment Screens have been developed, and research supports their ability to identify youth with emergent risk, and screen for youth who should receive more comprehensive mental health assessments. A study by Wasserman et al. (2004) confirms that such brief tools as the MAYSI-2 are useful for identifying youth who have a possible mental health problem so that they can be further evaluated by such tools as the DISC to identify a more specific diagnosis. These brief assessment tools should be utilized for these purposes and they are most effective

when administered promptly upon the youth's arrival at a secure setting.

Several of the assessments described in [Table 2](#) address the aim of tapping different domains of functioning in the assessment. Some tools utilize multiple sources of data such as files and collateral sources of information in addition to self report, while others do not. This reflects the constant struggle to provide thorough assessment with limited time/expense resources. Considering the importance of accurately assessing and treating offender mental health problems, we conclude that assessments that tap into a range of information sources are worth the time and effort.

It is unclear from the current empirical evidence how effective the above reviewed assessment tools are in re-assessing risk once the youth has been reintegrated into the community. Further research is needed to clarify whether current assessment tools are useful for post-incarceration re-assessment or whether other assessments, which take into consideration the importance of environmental transition, should be developed for that purpose.

It is important to mention that, while great progress has been made in beginning to understand and assess juvenile offender mental health risk, there is much work to be done in testing the ability of these assessment tools to generalize beyond the population for which they were developed. Recent research by Wasserman, McReynolds, Ko, Katz, and Schwank (2005) examining the prevalence of psychiatric disorders among youths at probation intake, reported that violent female offenders were up to five times more likely to report anxiety disorders than their male counterparts. Furthermore, of youth with conduct disorder, girls seemed to be more likely than boys to have complex diagnoses due to elevated rates of co-occurring internalizing disorders. Research also shows that ethnic minority youth are overrepresented in the juvenile justice system, yet few mental health risk assessment tools have been tested across gender or ethnic groups (Devine, Coolbaugh, & Jenkins, 1998). Further research is needed to evaluate assessment instruments with female offenders and ethnic minority offenders, as research suggests adaptations may need to be made to accurately assess the needs of these vulnerable groups (Cauffman et al., 2006).

One additional facet of risk assessment appears particularly lacking in the field of juvenile justice. Risk assessments of juvenile offenders need to identify those youths likely to re-offend into adulthood, and who are likely to be the chronic career criminals. Several classic studies have documented the pattern of desistance of delinquent behavior in young adulthood (Farrington & West, 1977; Gottfredson & Hirschi, 1990; Elliott, et al., 1983). Gottfredson and Hirschi's (1990) key finding was that most juveniles discontinue their delinquent acts in early adulthood. Farrington and West's classic study indicated that only 22.6 percent of their research subjects had subsequent convictions as adults. Elliott and associates found that only 2 to 3 out of every 10 adjudicated violent juveniles were arrested for violent crimes in adulthood. Also noteworthy is the fact that while the majority of delinquent youth do not seem to present a long-term risk of re-offending in adulthood, there is a small group of 5 to 6 percent of different cohorts that chronically persist in crime into adulthood and are responsible for a high volume of multiple offenses. Wolfgang, Figlio and Sellin (1972) reported in their birth cohort study that approximately 6 percent of their subjects were responsible for over 50 percent of the official crimes by the cohort. These chronic offenders were likely to have poor school grades and achievement, low IQ test scores, be of non-white racial background, low socioeconomic status, and school dropouts. Farrington (1985) found similar results of chronic offending by a small percentage of offenders—6 percent committing 49 percent of the criminal offenses.

Risk assessment instruments to date have not been well tested in their ability to differentiate those youth who will chronically offend into adulthood from those who are temporary adolescent offenders. Perhaps Hare's Psychopathy checklist comes closest to beginning to identify this type of particularly serious long-term offender. However, it is clear that, while short-term recidivism is important to assess, much work is needed to expand current mental health assessment tools to better identify the potential long-term chronic offenders.

[back to top](#)

References

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

**Table 1:
Strengths and Limitations of Commonly Used Brief Assessment
Screens**

Brief Assessment Screens				
Instrument (developer)	Subscales	Classification	Strengths	Limitations
mAYSI-2 Grisso and barnum (2003)	Alcohol/drug use, angry-irritable, depressedanxious, somatic complaints, suicide ideation, thought disturbance, and traumatic experiences	Youth self report	Low cost Brief Administration Time Ease of Administration	Potential social desirability Need thought disturbance for girls Test applicability to ethnic minority youth
pOSIT rahdert (1991)	Substance use/abuse, physical health, mental health, family relations, peer relations, educational status, vocational status, social skills, leisure/recreation, and aggressive behavior/delinquency	Youth self report	Identifies youth in need of further assessment Pblic domain instrument	Administer to one youth at a time Test applicability to ethnic minority youth
CAFAS Hodges (2000a)	School/work, home, community, behavior toward others, moods/emotions, self-harmful behavior, substance use, thinking, material needs and family/social	Parent rating Teacher rating Youth self report Structured observation	Easy training for administration Helps prioritize interventions Objective measures of functioning	Requires time investment in observing behaviors and collecting collateral information

Table 2: Strengths and Limitations of Commonly Used Comprehensive Assessment Tools

Comprehensive Assessment Screens				
Instrument (developer)	Subscales	Classification	Strengths	Limitations
<p>DISC</p> <p>Shafer, Fisher, Lucas, Dulcan, & Scwab-Stone (2000)</p>	<p>Anxiety, mood, disruptive behavior, substance use, and miscellaneous (eating disorders, tic disorders, etc.)</p>	<p>Youth self-administered, computerized, structured interview</p>	<p>Results in diagnosis allowing for more thorough planning</p> <p>No professional training required in administration</p> <p>Computer administration may ease discloser of suicidal ideation</p>	<p>Computer skills necessary</p> <p>Does not address other social or environmental domains</p> <p>Potential social desirability</p>
<p>mmpI-A</p> <p>Archer (1997)</p>	<p>Validity scales (e.g. defensiveness, tendency to exaggerate, response consistency), clinical scales (e.g. psychopathology such as depression, anxiety, schizophrenia, antisocial behaviors), content scales (e.g. externalizing behaviors, anger, low self-esteem), and supplementary scales (immaturity, repression)</p>	<p>Youth self report</p>	<p>Widely used</p> <p>Useful in assessing change over time</p> <p>Ease of administration</p>	<p>Requires trained professional to administer</p> <p>Ability to predict violent recidivism has not been evaluated</p>
<p>mACI</p> <p>millon (1993)</p>	<p>Validity scales (disclosure, desirability, debasement), reliability scale, clinical syndrome scales (eating dysfunction, substance abuse, delinquent predisposition, impulsive propensity, anxious feelings, depressive affect, suicidal tendency), personality scales (introversive, inhibited, doleful, submissive, dramatizing, egotistic, unruly, forceful, conforming, oppositional, self-demeaning, borderline tendencies), and expressed concern scales (identity confusion, self-devaluation, body disapproval, sexual discomfort, peer insecurity, social intensively, family discord, child abuse)</p>	<p>Youth self report</p>	<p>Minimum training for administrators</p> <p>Built in measure of validity and reliability</p> <p>Consistent with DSm-IV</p>	<p>Relies on client retrospective reports rather than file data</p> <p>More research needed to assess predictive ability in juvenile justice setting</p>

Table 3: Strengths and Limitations of Commonly Used Assessments of Future Risk

Assessments of Future Risk				
Instrument (developer)	Subscales	Classification	Strengths	Limitations
YLS/CmI Hoge, Andrews, & Leschied (2002)	Assessment of risk and needs (prior/current offenses, family circumstances, education/employment, peer associations, substance abuse, leisure, personality/behavior, and attitude); Summary of risk/need factors (comparing scales to normative ranges); Other needs/special circumstances (situational information such as parental drug use or behavioral records that add information specific to youth); professional override feature (asks clinician to use clinical judgment considering all relevant information to rate youth's risk level); Contact level (intensity of services recommended); Case management plan (specific goals and objectives for reaching goals)	Trained professional completes structured assessment using information from youth interview, file review, and collateral sources	Can be administered by "front-line" staff Assesses risks and needs	Requires time to review collateral materials Low reliability on one subscale
SAVrY bartel, borum, & Forth (2000)	Protective factors (prosocial involvement, strong social support, strong attachments and bonds, positive attitude toward intervention and authority, strong commitment to school, resilient personality traits); risk factors including: historical (history of violence, of nonviolent offending, early initiation of violence, history of self harm, childhood exposure to maltreatment, parental criminality, early caregiver disruption, poor school achievement), individual (negative attitudes, risk taking/impulsivity, substance use difficulties, anger management problems, low empathy/remorse, ADHD difficulties, poor compliance, low interest/commitment to school) and social/environmental (peer delinquency, peer rejection, stress/poor coping, poor parental management, lack of personal/ social support, community disorganization)	Examiner uses information from a systematic assessment of risk and protective factors collected through interview with youth and review of records (police/ probation, mental health, social service reports) to make a structured professional judgment	Does not provide a decision or cut off point requiring knowledge of how identified factors relate to behaviors	No formalized training provided Requires qualified examiners Predicts case specific violence not general violence likelihood
pCL:YV Forth, Kosson, & Hare (2003)	Impression management, grandiose sense of worth, stimulation seeking, pathological lying, manipulation of personal gain, lack of remorse, shallow affect, lack of empathy, parasitic orientation, poor anger control, impersonal sexual behavior, early behavior problems, lacks goals, impulsivity, irresponsibility, failure to accept responsibility, unstable interpersonal relationships, serious criminal behavior, serious violence of conditional release, and criminal versatility	Examiner uses information from an intensive interview with the youth, collateral sources, and review of the chart to rate the youth on 20-item checklist	Identifies risk factors for potentially very serious offenders	Complex training and advanced graduate degree recommended for administering assessment Controversy over stigmatizing youth with psychopathy label

[Home](#)

Mental Illness in Correctional Populations: The Use of Standardized Screening Tools for Further Evaluation or Treatment

[References](#)

Arthur J. Lurigio
Departments of Criminal Justice and Psychology
Loyola University Chicago

James A. Swartz
Jane Addams College of Social Work
University of Illinois at Chicago

[Mentally Ill in Criminal Justice Settings
Need for Mental Health Screening Tools
Candidate Screening Instruments
SMI Screening Tools for Criminal Justice
Summary and Conclusions](#)

CRIMINAL JUSTICE AGENCIES have become de facto settings for mental health treatment and other clinical services (Lurigio & Swartz, 2000). The growing number of persons with serious mental illness (SMI) who appear at each step in the criminal justice process—from arrest to post-incarceration release—has forced professionals in corrections, who usually have no backgrounds or experiences in the mental health field, to continually face the challenges of identifying, referring, and case managing the mentally ill. With or without special training or guidelines, most criminal justice professionals (such as police officers, probation officers, prison intake workers, and parole agents) screen individuals for mental illness in their daily practice. They do so in order to make decisions about such options as diversion, segregated housing, treatment, or other specialized interventions.

Screening for SMI involves a brief initial evaluation about a client's need for mental health services, which can be done at the point of arrest, sentencing, or imprisonment. It also can be done formally or informally, and can trigger either an immediate decision or a more comprehensive psychiatric evaluation designed to help make subsequent mental health-related decisions about a case. For example, police officers at the scene determine who should be diverted for an emergency hospitalization instead of arrested and who is at risk of attempting suicide in lockup. Probation officers conduct mental health screening as part of an overall needs assessment to determine offender classification and service brokerage. Detention and correctional officers screen incoming prison inmates and jail detainees for mental illness in order to assign them to specialized housing and programming.

This article examines the use of actuarial screening tools that have been developed to flag persons with SMI for further assessment, diagnosis, and treatment in institutional and

community-based correctional facilities. In the absence of such tools, the mentally ill in the criminal justice system are likely to go unrecognized and untreated. The paper is divided into four major sections. The first discusses the prevalence of persons with SMI in jail, prison, and probation populations and the dearth of mental health services for them. The second emphasizes the use of valid and reliable screening tools as an important first step in providing services for offenders with mental illness. The third presents the results of two studies that have tested mental health screening tools for use with criminal justice populations. The fourth concludes with recommendations for the incorporation of mental health screening tools in the intake protocols of correctional departments.

[back to top](#)

Mentally Ill in Criminal Justice Settings

Over the past 20 years, several epidemiological studies have shown that substantial numbers of persons involved in the criminal justice system have SMI, such as schizophrenia, bipolar disorder, and major depression (Abram & Teplin, 1991; Abram et al., 2003; Ditton, 1999; Diamond et al., 2001; Human Rights Watch [HRW], 2003). The three largest psychiatric facilities in the United States are urban jails: the Los Angeles County Jail, the Cook County Jail (CCJ) in Chicago, and the jail at Riker's Island in New York City (Insel, 2003). One estimate suggests that 900,000 individuals with SMI are admitted to our nation's jails annually (Steadman, Scott, Osher, Agnese, & Robbins, 2005). Many factors explain the large numbers of the mentally ill in offender populations. These factors include transinstitutionalization, stricter civil commitment laws, homelessness, public order policing, and the fragmentation of the mental health and drug treatment service systems (HRW, 2003; Lamb & Weinberger, 1998; Lurigio, 2005; Lurigio & Swartz, 2000).

Ditton (1999) has conducted the only national study to date on the prevalence of the mentally ill in correctional populations. She reported that, at midyear 1998, an estimated 283,800 mentally ill offenders were incarcerated in our nation's prisons and jails. A total of 16 percent of those surveyed in each population reported either a mental health condition or an overnight stay in a mental hospital. Approximately 16 percent, or an estimated 547,800 probationers, indicated that they had experienced in their lifetime a mental disorder or stayed overnight in a mental hospital.

Based on information from personal interviews, state prison inmates with a mental disorder were more likely than other inmates to be incarcerated for a violent offense (53 percent compared with 46 percent) and to be under the influence of alcohol or drugs at the time of the current offense (59 percent compared with 51 percent). They also were more than twice as likely as other inmates to have been homeless in the 12 months before their arrest (20 percent compared with 9 percent). More than three-quarters of mentally ill inmates had been sentenced to prison, jail, or probation at least once prior to their current sentence. Since admission, 61 percent of mentally ill inmates in state prison and 41 percent of mentally ill detainees in local jails reported that they had received treatment for a mental health problem, including counseling, medication, or other mental health services (Ditton, 1999).

The large number of mentally ill offenders underscores the need for effective mental health screening and treatment services. Courts have ruled consistently that jails and prisons are legally obligated to provide mental health and other medical services to detainees and prisoners (Diamond et al., 2001; Veysey & Bichler-Robertson, 2002). Left untreated, incarcerated persons with SMI have trouble adapting to life in prisons and jails, as well as following the written and unwritten rules that are inherent in the daily routines of correctional facilities (HRW, 2003; Torrey, 1995). Furthermore, prisoners and detainees with SMI are at high risk for suicide, disciplinary infractions, and victimization (Dicataldo et al., 1995; HRW, 2003). Similarly, mentally ill probationers often have trouble complying with their probation orders, such as reporting to their probation officers or finding employment and housing opportunities. These difficulties increase their risk for a technical violation or new arrest (Solomon & Draine, 1999).

The landmark case of *Ruiz v. Estelle* (503 P. Supp. 1265.1323 [1980]) set forth standards for

“minimally adequate mental health treatment programs” in prisons. These standards consist of the systematic mental health screening and evaluation of inmates; the capacity to ensure that treatment involves more than just inmate segregation; the provision of individualized treatment by trained mental health professionals; the maintenance of accurate and complete mental health records; the supervision and review of prescriptions; and the identification of inmates with suicidal tendencies (Jemelka et al., 1993). Other case precedents have established jail detainees’ rights to mental health treatment and aftercare services (e.g., *Brad H. et al. v. City of New York et al.* . 185 Misc. 2d 420; 712 N.Y.S. 2d 336 [Sup. Ct. 2000]).

Notwithstanding the clear legal mandate to provide mental health services, and the prodigious numbers of offenders who need and receive such services, many mentally ill individuals remain unidentified and untreated while under the jurisdiction of the criminal justice system (Elliot, 1997; National Commission on Correctional Health Care [NCCHC], 2002; Steadman & Veysey, 1997; Teplin, 1990). For example, in a study of female jail detainees, Teplin et al. (1997) found that only 25 percent of those meeting the criteria for SMI received treatment within one week of admission. As Ditton (1999) found, nearly 40 percent of prison inmates and 60 percent of jail detainees with mental illness reported that they were receiving no mental health services during their recent incarceration.

These findings mirror those of broader studies of mental health care for prisoners (ACP et al., 1992; Jordan et al., 1992) and probationers (Lurigio et al., 2003), which also have found high rates of untreated mental illness. For example, a survey of probation departments by Boone (1995) found that only 15 percent of those responding to a national survey had programs for the mentally ill. Several factors contribute to the criminal justice system’s failure to deliver adequate treatment to individuals with SMI. These include the scarcity of mental health resources, the rapid turnover of detainees in jails, correctional staff with little clinical training, and significant increases in the criminal justice population, which exceeds the capacity of criminal justice organizations to implement mental health services for all persons who require such care (NCCHC, 2002).

[back to top](#)

Need for Mental Health Screening Tools

Clinicians use several techniques at intake to collect information for formal psychiatric evaluations. In general, they examine the type and severity of the symptoms (what the patient reports) and signs (what the clinician observes) of SMI. The two key components of the evaluation process are the construction of treatment histories, including the use of medications, and the performance of mental status examinations to evaluate current levels of cognitive, social, behavioral, and emotional functioning. Information gathering techniques include structured and unstructured interviews, symptom questionnaires, patient observations, personality inventories, and psychological tests. The clinician integrates the findings from these various sources in order to render a diagnosis and recommend a psychiatric treatment and rehabilitation plan (Nolen-Hoeksema, 2005).

The basic psychometric properties of psychiatric assessment tools are validity and reliability. The former refers to the accuracy of instrumentation and the latter to the consistency with which such instrumentation is used to collect information. The field of psychometrics has established different types of validity and reliability for different purposes of instrument construction and application (Anastasi & Urbina, 1997). Two other concepts useful in determining the accuracy of an assessment or screening tool are sensitivity and specificity. Sensitivity is the likelihood that a tool will find disease among those who have the disease, or the proportion of people with disease who have a positive test result. Specificity is the likelihood that a tool will find no disease among those who do not have the disease, or the proportion of people free of a disease who have a negative test result.

As we suggested in our preceding discussion, psychiatric assessment requires considerable time, skill, and training in the use of specialized diagnostic tools and criteria. Few correctional staff

persons have such backgrounds or knowledge. Clinicians often are hired in correctional settings to perform assessments and treatment, but their time is usually limited and their caseloads are overwhelmingly high (Earley, 2006). Hence, it is infeasible and costly to conduct a full-blown psychiatric assessment on every person entering the criminal justice system.

To conserve sparse clinical time and resources, mental health screening should be employed; its goal is the identification of persons who need further evaluation. However, if mental health screening is done at all in criminal justice settings, screening tools are usually unstandardized, and screening procedures vary greatly within and among agencies. The validity and reliability of such screening results are quite low (Steadman & Veysey, 1997). For example, a survey of state prisons found most of the responding institutions had no valid data on the prevalence of various psychiatric conditions (primarily SMI) in their inmate populations—a problem attributable to antiquated information systems as well as to non-standardized screening and assessment procedures (Hornung et al., 2002).

Although no comparable surveys have been conducted of probation departments or jails, a similar degree of variation in screening and referral practices is likely to be found in those settings as well (Lurigio et al., 2003; Skeem et al., 2003). As Steadman, et al. (2005) noted about screening for mental health problems in jails: “Screening may consist of anything from one or two questions about previous treatment to a detailed, structured mental health status examination” (p. 816).

Overall, few criminal justice agencies employ brief screening tools that are constructed and validated specifically for offender populations (Swartz, 2001). Since the inception of the war on drugs, criminal justice administrators and practitioners have encouraged the use of screening and assessment tools to detect substance use disorders. However, they have paid little or no attention to the development of comparable instruments for detecting mental illness (Peters et al., 2000).

Many criminal justice agencies construct their own psychiatric screening tools, which rarely are subjected to rigorous reliability and validity studies and often are based simply on face validity, which is the lowest level of measurement accuracy. Face validity requires only that a tool appears to be measuring what it purports to measure. Still others rely on probation officers’ subjective judgments about mental health needs, which tend to grossly underestimate the number of probationers with SMI (Lurigio & Swartz, 2006). The best screening tools have high predictive validity: “most of the people who are flagged by [the tool] as being positive should, on assessment, be found to have a treatable serious mental illness” (Steadman, et al. 2005, p. 817).

The screening instrument presently used in the CCJ is typical. More than 300 detainees are processed through reception and classification every day at the jail. They are screened for psychiatric problems by mental health specialists who ask a short series of questions about previous psychiatric hospitalizations, current use of psychiatric medication, and recent thoughts of suicide. This set of questions is consistent with recommended screening practices for jails and prisons and is effective in identifying many individuals who require further assessment and treatment (American Psychiatric Association [APA], 1989; Steadman & Veysey, 1997). Nonetheless, the questions are likely to miss numerous other individuals with SMI, specifically, those who have no previous hospitalizations or current suicidal ideation, or who are taking no psychiatric medications (Teplin, 1990; Teplin et al., 1997; Teplin & Swartz, 1989). Teplin (1990) reported that these criteria missed nearly two-thirds of the mentally ill detainees who were identified with an independently administered and standardized clinical instrument.

As we noted above, screening instruments for psychiatric disorders have been validated on samples from the general population but not on samples from criminal justice populations. The accurate estimate of the underlying prevalence of a disorder is important for clinical decision-making purposes; for example, a determination of when a person’s symptoms suggest that further evaluation and treatment are warranted (Schmitz et al., 2000). Thus, a screening tool for correctional settings should be tested on criminal justice populations, where the prevalence of psychiatric disorders is higher than it is in the general population (HRW, 2003). The choice of a

corrections-based screening tool also must consider the conditions that apply to screening in criminal justice settings, which place constraints on the kinds of screening instruments that are best suited for those contexts where there is a shortage of clinical services and treatment slots and a high volume of screenings, which are conducted by correctional staff persons who have no clinical or diagnostic expertise.

Hepburn (1994) recommended that screening instruments for substance use disorders in criminal justice settings have standardized and replicable scoring criteria that can be implemented and interpreted by lay interviewers. They also should be brief and easy to administer without the need for extensive training. Screening for psychiatric illnesses in a criminal justice setting is subject to the same contextual limitations and pressures as screening for substance use disorders. Therefore, the ideal psychiatric screening device would have the same properties as a substance abuse screening device.

Defining the appropriate content of a screening instrument for psychiatric disorders is much more difficult than it is for substance use disorders, for two basic reasons. First, the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR), Axis I (Clinical Disorders) has 15 general categories of non-substance use disorders (APA, 2002). It is possible to screen briefly for all drugs of abuse or simply for the presence of any substance use problem. However, it is impossible (at the very least, highly unwieldy) to efficiently screen for every Axis I psychiatric diagnosis. Second, even if it was possible to screen for every DSM-IV, Axis I disorder, not everyone with a disorder needs treatment. Clinical severity and treatment need lie on a continuum; even some persons with seemingly severe psychiatric disorders are able to function adequately without clinical intervention (Regier et al., 1998). The challenge is to determine which psychiatric disorders should be included in a screening tool and to define when a disorder is severe enough to warrant clinical intervention or further assessment.

[back to top](#)

Candidate Screening Instruments

A couple of approaches have been adopted to simplify standardized screening procedures for psychiatric disorders. The first determines if an individual meets the diagnostic criteria for a few diagnoses that are likely to be clinically severe and to require treatment and other interventions. This restricted subset of all DSM-IV diagnoses typically includes those that are widely regarded as SMIs: schizophrenia, bipolar disorder, and major depressive disorder. Similar to the longer assessment instruments for psychiatric diagnosis, such as the Composite International Diagnostic Interview (CIDI) (Robins et al., 1988) from which they were derived, these instruments are modular. Each module consists of a sequence of questions for diagnosing a specific disorder or class of disorders.

Among the tools that adopt a diagnostic approach to screening are the Composite International Diagnostic Interview – Short Form (CIDI-SF) (Kessler et al., 1998), the Mini-Neuropsychiatric Interview (MINI) (Sheehan et al., 1998), and the Referral Decision Scale (RDS) (Teplin & Swartz, 1989). The administration time for these relatively brief instruments can be shortened further by omitting modules that screen for disorders that are of little interest to clinicians or researchers.

Despite their administrative flexibility, problems of over- and under-identification with this class of instruments limit their usefulness, particularly in resource-constrained criminal justice settings in which accuracy is crucial. The diagnostic approach to screening equates the need for clinical intervention with diagnosis. Persons who meet the diagnostic criteria for one or more disorders are referred for further assessment and possibly treatment. Those who meet none of the full criteria for any disorders are not referred for either further assessment or treatment.

The potential drawback of such techniques is that they miss persons who have a severe disorder that is not contained in the screening tool (such as post-traumatic stress disorder, generalized anxiety disorder) but that nevertheless requires clinical intervention. These errors are called false

negatives: concluding that a person with mental illness has no disorder. Such individuals might receive no treatment, while detained or under supervision, despite the clinical severity of their conditions. They are at risk for problems associated with untreated SMI (e.g., rearrest, homelessness, violence, substance abuse). Moreover, despite the administrative flexibility of being able to select the diagnoses that are included in the screening, the necessity of obtaining a DSM-IV diagnosis adds a level of complexity to the instruments that is attributable to the use of skip patterns and question probes. Given the lack of clinical and interviewing skills among corrections staff, the inclusion of even a small number of skip patterns and probes can sharply reduce the validity and reliability of the instrument.

The problems related to using a diagnostic approach to screening for psychiatric treatment have recently led to a second approach that de-emphasizes diagnosis and focuses instead on symptom severity and level of impairment (Kessler et al., 2002). Although this approach has been discussed widely in the literature (see Murphy, 2002), it has recently gained currency because large-scale epidemiological surveys, such as the Epidemiological Catchment Area (ECA) study and the National Comorbidity Survey (NCS), have found surprisingly high prevalence rates of psychiatric disorders (Kessler et al., 1994; Regier et al., 1990).

In both studies, between 20 and 30 percent of the general population met the DSM-IV criteria for at least one past-year Axis I disorder. As it is unlikely that this large a proportion of the general population required mental health treatment services, the findings of these surveys were of limited usefulness in guiding federal and state treatment resource allocations. These findings also suggested that screening for symptom severity and level of functional impairment is a more efficient way of discerning the need for psychiatric treatment (see Regier et al., 1998; Slade & Andrews, 2002).

The tools related to the symptom-severity approach are particularly applicable for use in criminal justice settings. Their advantages include the use of briefer screening instruments, without skip patterns and probes, which makes the screening tool simpler to administer. Such screening tools can be implemented by lay interviewers to identify individuals with the most severe psychiatric disorders, regardless of diagnosis. This approach conserves limited resources for only those mentally ill persons most in need of services. In other words, such tools can avoid false positives, which identify as mentally ill persons whose symptoms are not severe enough to warrant treatment. A low false positive rate is especially important in criminal justice settings, in which scarce mental health resources must be used sparingly (Steadman et al., 2005).

[back to top](#)

SMI Screening Tools for Criminal Justice

K6/K10. Among the class of instruments that take this approach to screening are the K6/K10 scales that, for the reasons we mentioned above, appear to be especially promising for use with criminal justice populations (Kessler et al., 2002). Kessler et al. (2002) began the tool-building process with a large pool of items derived from an extensive battery of psychological instruments. They then used analytic procedures, based on item-response theory, to distill a subset of 10 questions (the K10) and a completely overlapping subset of 6 questions (the K6).

The K6 asks respondents how often in the previous month they felt “nervous,” “so sad that nothing could cheer [them] up,” “restless or fidgety,” “hopeless,” “everything was an effort,” and “worthless.” These questions identified with maximum sensitivity the individuals who met the following two criteria: a past-year diagnosis of any DSMIV Axis I psychiatric disorder and a Global Assessment of Functioning (GAF) score below 60 (i.e., moderate to severe impairment in functioning) (see APA, 2002, and Endicott et al., 1976).

Further calibration of the K6/K10 scales was done to select cutoff scores that identify individuals above the 90 th percentile in symptom severity, consistent with estimates which suggest that 6-10 percent of the general population needs psychiatric treatment services at any given time (Kessler et al., 2002). In validity studies, the K6 scale has performed as well as the K10 in identifying

individuals with SMI and has become the more widely used instrument (Kessler et al., 2003). The K6 is now included in national surveys, such as the National Survey on Drug Use and Health (NSDUH) and the National Health Interview Survey (NHIS).

Despite its widespread application in general population studies of psychiatric disorders, we are the only researchers to validate the K6 for use with a criminally involved sample of persons. We demonstrated the use of the K6 scale with a sample of adults who reported an arrest in the past year, and compared the classification results of the K6 screen with those obtained using a common, but unvalidated, set of screening questions (e.g., receipt of past psychiatric treatment services, use of prescribed psychiatric medications). Specifically, we compared the diagnostic accuracy of the unvalidated set of questions with the K6 classification results, and examined the characteristics of participants who are misclassified in order to understand why they were incorrectly identified by standard criteria. We also examined the characteristics of offenders who screened positive on the K-6 scale for SMI, compared with those who screened negative (Swartz & Lurigio, 2005; Swartz & Lurigio, in press).

We found that nearly 20 percent of the 1,700 participants with a past-year arrest in the 2002 NSDUH sample had a K6 score of 13 or higher, indicating that in the past year they had experienced symptoms of severe psychological distress consistent with the presence of SMI. Our findings also indicated that all the items in the K6 work equally well in detecting SMI among arrestees for both genders. In addition, we found the same pattern of item and scale consistency among different racial/ethnic and age groups. Respondents with SMI were more likely than those with no SMI to report a past-year substance use disorder that met the DSM-IV criteria for dependence or abuse; they also were more likely to have received drug abuse and mental health treatment in the past year (Swartz & Lurigio, 2005; Swartz & Lurigio, in press).

Brief Jail Mental Health Screen. The Brief Jail Mental Health Screen (BJMHS) was derived from the RDS (Teplin & Swartz, 1989) and consists of eight dichotomous questions. The first six ask respondents whether they currently believe someone is putting thoughts into, or taking them out of, their heads; whether they currently believe that other people read their minds; whether family or friends have noticed that they are more active than usual; whether they have gained or lost weight for several weeks without trying to do so; and whether they have currently felt useless or sinful. The last two questions ask whether they are currently taking any medication for emotional or mental health problems and whether they have ever been in a hospital for emotional or mental health problems.

To test the concurrent validity of the BJMHS, Steadman et al. (2005) studied samples of detainees who were and were not referred for mental health services based on the scale scores. Each of the detainees in those groups also was evaluated with the Structured Clinical Interview for DSM-IV (SCID), which provides a diagnosis-driven assessment of mental health problems. In this study, the SCID served as the gold standard, or the accepted reference or diagnostic test, for psychiatric illness. Results showed that the BJMHS correctly classified as having a diagnosable mental illness nearly 75 percent of the male detainees but only 62 percent of the female detainees. The researchers concluded that the BJMHS is, overall, a practical, simple, and efficient tool for psychiatric screening in jails, but that it has an unacceptably high false-negative rate for women.

[back to top](#)

Summary and Conclusions

Individuals with SMI are over-represented in the criminal justice system. Since Abramson's (1972) seminal work on the criminalization of the mentally ill, such persons have become more abundant in terms of their absolute numbers and proportionate representation in correctional populations. Their prevalence in jails, prisons, and probation caseloads is likely to grow even further unless fundamental changes take place in the circumstances that contribute to their involvement in the criminal justice system. However, the historical factors that encouraged the criminalization of the mentally ill are still common today: the lack of community-based mental

health services for persons with SMI, the splintering of the drug abuse and mental health treatment systems, and the legal restrictions on involuntary commitment (Lurigio, 2005).

Persons with SMI often reside in environments where crime and opportunities to commit crime are rampant, drug and crime enforcement strategies are aggressive, and mental health and drug abuse treatment programs are limited or inaccessible—all of which increase the likelihood that the mentally ill will be caught in the criminal-justice web (Draine, 1993; Fisher, Silver, & Wolff, in press). In addition, the co-occurrence of psychiatric and substance use disorders, which is more common in the criminal justice population than in the general population, elevates the risk of arrest, violent behavior, hospitalization, incarceration, recidivism, and a host of other adverse life events. The failure to treat SMI, substance use disorders, and their comorbidity jeopardizes public safety, promotes recidivism, and can result in legal liability for criminal justice departments that are unresponsive to clients' needs for mental health services (Lurigio, 2003).

Effective screening is the first step in properly addressing the behavioral healthcare problems of the mentally ill in the criminal justice system. Mental health resources within agencies and communities are scanty; therefore, accurate identification for direct service provision or referral is paramount. As we have noted in this article, corrections professionals typically have no background, training, or experience in the assessment or treatment of mental illness. Moreover, information on treatment history and current use of psychiatric medication is not always an accurate indicator of a current mental health problem. Similarly, subjective judgments regarding mental health treatment need are fraught with errors and lead to inconsistent and biased decisions among probation officers and other professionals, including those with expertise in the mental health field (Lurigio & Swartz, 2006).

We recommend the use of the K-6 as a promising screening instrument for criminal justice populations. The BJMHS is also a viable option for such purposes but should be further refined to improve its specificity with women offenders. For both tools, more validation studies should be conducted to test their preciseness with different correctional populations. In addition, the construction of mental health screening tools for correctional clients should focus on identifying persons with comorbid psychiatric and substance use disorders. Such a screening tool is presently being developed as part of the large-scale, federally funded project known as the Criminal Justice-Drug Abuse Studies Initiative (CJ-DATS), which is funded by the National Institute on Drug Abuse (Sacks, Melnick, & Coen, 2005). Finally, feasibility, time, and resource studies must be undertaken to examine the effects of widespread mental health screening in criminal justice settings. The bottom line is whether screening actually results in better outcomes for individuals with SMI who are involved in the criminal justice system.

[back to top](#)

[References](#)

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

[Home](#)

Risk Assessment and Sex Offender Community Supervision: A Context-Specific Framework

[References](#)

Andrew J. Harris Ph.D.

University of Massachusetts, Lowell

[Sexual Offender Risk Assessment—The State of the Field](#)
[Risk Assessment and Community Supervision Practice: A Framework](#)
[Conclusion](#)

IN BOTH RESEARCH and practice, the past two decades have produced considerable developments in the field of specialized risk assessment for sexual offenders. Dozens of studies have contributed to a growing evidence base regarding recidivism risk factors and the potential efficacy of treatment interventions (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2004). On the basis of these findings, a broad array of specialized actuarial and guided clinical assessment instruments have been introduced and continue to be tested and refined (Doren, 2004a).

These developments have been fueled in part by resurgent policy attention to the issue of sex offender management during the past 15 years. Prompted in part by federal legislation, registration and community notification laws have been adopted nationwide, calling for effective systems of classifying levels of risk (Adams, 2002). Since 1990, seventeen states have adopted civil commitment laws for sexual predators—policies predicated on predictions of future violence and increasingly requiring states to address the complex issues associated with an individual's suitability for community release (Harris, 2005). The emergence of specialized models for community supervision of sexual offenders has demanded effective means of applying risk assessment in a multi-disciplinary context (English, Pullen, & Jones, 1997). The introduction of risk-based sentencing systems has produced unique demands for evidence-based decision tools that can both inform the sentencing process while ensuring due process (Kern & Farrar-Owens, 2004). As of this writing, legislative activity shows few signs of slowing, with the issue of sex offenders remaining at the top of state legislative crime control agendas (National Conference of State Legislatures, 2006).

While increased policy focus on these issues has produced both a significant expansion of the evidence base and the increased attention of researchers, it has also presented the burgeoning field of sex offender risk assessment with considerable challenges. The often overlooked heterogeneity of the population to be managed, coupled with the diverse range of organizational and programmatic contexts in which risk assessment is required, produces considerable potential for over-generalization of particular findings or the misapplication of particular tools or approaches.

As our policies have evolved, sex offender risk assessment has been called upon to respond to the needs of multiple stakeholders and to meet a wide range of legal, forensic, and clinical

purposes. Treatment professionals use it to develop treatment plans or evaluate progress. Probation and parole personnel use it to establish suitability for community supervision, case management, and intervention. The courts apply it for purposes of civil commitment or criminal sentencing. Law enforcement has adopted it for purposes of profiling, investigation, or designation of sex offender risk levels for purposes of registration and community notification.

Considering this diversity of contexts, it would appear that one-dimensional “debates” over the relative merits of particular approaches to risk assessment (e.g. clinical vs. actuarial approaches) may miss an essential part of the picture. Indeed, the key to finding “middle ground” between competing viewpoints may rest primarily in achieving greater clarity of our goals and objectives, and in adapting our methods and practice to meet those goals (Dvoskin & Heilbrun, 2001).

Consistent with this view, this article aims to present the discussion regarding sexual offender risk assessment in a circumscribed programmatic context, with specific focus on the practice of community-based supervision of sexual offenders.

This article consists of two parts. The first reviews the current state of sex offender risk assessment, considering the factors known to be associated with sexual recidivism and the methods currently utilized to translate those factors into risk assessment practice. The article’s second section applies this understanding to the specific programmatic context of community-based supervision of sexual offenders, and sets forth a framework for integrating current risk assessment knowledge into systems of community-based supervision of sexual offenders.

[back to top](#)

Sexual Offender Risk Assessment—The State of the Field

As noted in the introduction, advances in the field of specialized sex offender risk assessment accelerated greatly beginning in the 1990s. While this development was spurred in part by advances in general violence risk assessment research, it was greatly facilitated by a range of critical developments in the realm of public policy surrounding sexual offenders (Hanson, 2005). On the legislative front, these developments included the spread of sex offender registration and community notification laws across the nation and the passage and implementation of civil commitment laws for sexually violent predators. Concurrently, with growing national emphasis on issues surrounding prisoner re-entry and community corrections, jurisdictions across the country were expanding and refining a range of specialized community supervision models for sexual offenders (English et al., 1997).

The common thread running through each of these diverse policy strategies is the shared goal of reducing recidivism among individuals previously convicted of sexual offenses. Consistent with this goal, policymakers, researchers, and practitioners have focused increasing attention on three main areas:

- The identification of risk factors associated with sexual recidivism;
- The integration of those risk factors into structured assessment instruments; and
- The refinement of interventions aimed at reducing re-offense rates.

This section addresses the first two areas, setting the stage for a later discussion of the implications of risk assessment for guiding community supervision interventions.

Risk Factors for Sexual Re-Offense

In 1998, Hanson and Bussiere published a meta-analysis of 61 studies providing information on 28,972 sexual offenders and investigating risk factors associated with sexual recidivism (Hanson & Bussiere, 1998). This was followed in 2004 by an updated metaanalysis encompassing the initial studies plus additional research conducted between 1998 and 2003, accounting for a total of 95 studies and over 31,000 sexual offenders (Hanson & Morton-Bourgon, 2004).

The studies found aggregate sexual reoffense rates (based on average follow-up periods of 5-6

years) of 13.4 percent and 13.7 percent, respectively. The earlier study further differentiated sexual re-offense rates between child molesters (12.7 percent re-offending) and rapists (18.9 percent reoffending). The studies also reported on rates of re-offense related to non-sexual violent crimes, finding overall re-offense rates of 12.2 percent and 14 percent (9.9 percent of child molesters and 22.1 percent of rapists).

Identified risk Factors

While the later study included some notable additional findings, the two analyses were fairly consistent in their overall conclusions regarding the major predictors of long-term recidivism.

Topping the list in both studies was the presence of certain types of sexual deviancy, as measured by both phallometric assessment and deviant sexual preferences as measured by standardized tools or clinical records. The Hanson and Bussiere study found sexual interest in children (i.e., pedophilia) to be a strong predictive factor in child molesters, although it did *not* find sexual interest in rape to be a significant predictor among rapists. The later study confirmed this finding, adding the existence of other paraphilias (such as exhibitionism and voyeurism) as having additional predictive value. The study further cautioned that the lack of findings regarding an association between sexual reoffense and paraphilic interest may be due to a limited number of studies that investigated this association, and suggested further research in this area.

The second most dominant factor identified in both studies involved the presence of antisocial lifestyle and orientation, as characterized by “rule violations, poor employment history, and reckless, impulsive behavior.” (Hanson & Morton-Bourgon, 2004). Notably, in contrast with the sexual deviance variables, this factor has been consistently found to serve as a strong predictor of general recidivism in non-sexual criminals (cite Bonta et al. 1998). Considering this, some have questioned whether similar mechanisms are at work in sexual and nonsexual offenders, or if antisocial orientation interacts with other predictors to create a unique dynamic among sexual offenders.

Beyond these first two major factors—sexual deviance and antisocial orientation—the meta-analyses identified a range of additional factors established as having moderate predictive value. These factors included:

- Age (younger offenders presenting higher risk);
- Number of prior offenses;
- Single marital status;
- Treatment failure;
- Sexual preoccupations; and
- Intimacy deficits.

Methodological Considerations

Despite the growing base of knowledge related to risk factors for sexual recidivism, research in this area has been constrained by a range of methodological issues. Most of these issues relate, in one way or another to the *base rate* of sexual offending—i.e. the proportion of individuals within the population who eventually re-offend.

As noted above, the aggregate base rate for sexual offender as established by studies included in meta-analyses is somewhere between 13 percent and 14 percent. For a variety of reasons, however, this figure most likely underestimates the “true” rate of sexual offending, and additionally does not effectively capture the range of variation in this rate across subsets of the sex offender population. Issues commonly associated with the base rate include:

Under-reporting — The vast majority of studies addressing the issue of sexual recidivism operationalize re-offense as incidents that are detected and lead to arrest and conviction. It is fairly well-established that only a limited proportion of sexual crimes—perhaps fewer than one in three—are reported to the police (Hart & Rennison, 2003). Accordingly, it is likely that actual re-offense rates may be substantially higher than those captured by recidivism researchers. A

related confounding factor relates to the fact that the extent of this under-estimation may not be uniform across groups of offenders, considering that offenders with certain characteristics (such as higher intelligence) may simply be more adept at avoiding detection.

Population heterogeneity —Sexual offenders are an extremely diverse group. Beyond the fundamental distinction between rapists and child molesters, each of these groups includes a wide range of subtypes linked to victim choices, underlying motivations, behavioral patterns, and other factors (Knight & Prentky, 1990; Lanning, 1986). This heterogeneity has a range of implications for both research and practice. From a methodological vantage, failure to effectively distinguish between these subgroups in research designs complicates the capacity to conduct within-group analysis, especially with those groups that are under-represented in samples or those with low overall base rates. In terms of application, this diversity of the offender population is generally not acknowledged in commonly used actuarial tools, leading some to question the validity of these instruments as means of predicting violence in individual cases (Hart, Webster, & Menzies, 1993).

Timeframes —For reasons of resources and practicality, studies employ a wide range of follow-up periods in their assessment of recidivism. Although the studies included in Hanson and Bussiere’s meta-analysis involved an average follow-up period of five years, evidence suggests that the risk of re-offense may extend far beyond this threshold (Hanson, Stey, & Gauthier, 1993). Hence, the research time horizon must be viewed as a source of potential bias in the derivation of the base rate.

Statistical Significance —From a research standpoint, the most immediate implication of a relatively low observable base rate involves researchers’ reduced capacity to draw statistically significant conclusions from available data. While this may be mitigated in part by increasing the sample size, many studies are limited in their capacity to expand their samples due to resource or logistical constraints.

Diminished Predictive Value —The predictive capacity of actuarial risk assessment instruments is directly influenced by the base rate upon which that instrument has been based and validated—the lower the base rate, the higher the probability of error. In general, low base rates are most likely to increase the probability of “false alarms.”

Static and Dynamic Factors

Reviewing the major variables known to be most closely associated with long-range sexual recidivism, many have noted that the majority of these factors are either static or highly stable in nature. While we have developed a fairly good sense of these immutable case characteristics that might place certain individuals at higher risk of re-offense than others, we have a much more limited understanding of the influence of dynamic characteristics associated with sexual recidivism risk (Craissati & Beech, 2003; Hanson & Harris, 2000b).

Certainly, the use of static variables in an applied context carries some clear advantages. Beyond their long-range predictive value, they are comparatively easily accessible through official records, and generally involve little or no subjective judgment. Yet over time, these static predictors gradually lose their utility for the majority of offenders under community supervision. Hence, effective systems of community supervision may begin with an understanding of an individual’s general risk as predicted by static variables, but ultimately depend on our capacity to identify and respond to changes in risk levels over time.

Dynamic variables may be viewed in terms of stable and acute factors (Gendreau, Little, & Goggin, 1996; Hanson & Harris, 2000b). *Stable factors* are those mutable characteristics of the individual that may change over time, but are not generally subject to short-term fluctuations. Key stable dynamic factors include variables such as cognitions, insight, treatment compliance, and attitudes related to offending behaviors. *Acute factors* reflect case characteristics that may change over more limited periods of times—in some cases weeks, days, or even hours. These factors might include both short-term life changes in domains such as employment, residence, or

relationships, and immediate conditions such as intoxication or circumstances that may provide access to potential victims.

Research regarding the effects of dynamic variables on sex offense recidivism remains in a developmental state. While methodological limitations have constrained much research in this area, key dynamic factors that appear to be related to recidivism include social adjustment, attitudes towards victims, self-awareness regarding risk to recidivate, victim access, and cooperation with supervision and treatment (Hanson & Harris, 2000b). The results of the Dynamic Supervision Project, a five-year longitudinal study currently following 1,000 offenders under community supervision in Canada, Alaska, and Iowa, may eventually provide further perspective on these factors (Harris & Hanson, 2003).

Approaches to Sex Offender risk Assessment

Hanson (2002) cites three potential approaches to sexual offender risk assessment— *pure actuarial* approaches, which make predictions based on survey instruments that leave no room for subjective interpretation; *guided clinical* approaches, which rely on the systematic professional judgment of qualified professionals based on empirically-derived instruments; and *adjusted actuarial* approaches in which professional judgment is superimposed on actuarial scores. To these options, we might add a fourth plausible approach—the use of *unstructured clinical judgment* to determine risk.

Doren (2004a) identifies over 20 instruments that have been applied in the assessment of sex offender risk. These instruments are varied—some have been developed as a means of evaluating the potential for general violence risk (Quinsey, Harris, Rice, & Cormier, 1998; Webster, Douglas, Eaves, & Hart, 1997), while others have been geared towards identifying the risk specifically for sexual offenders (Epperson et al., 1999; Hanson, 1997; Hanson & Harris, 2000a; Hanson & Thornton, 1999; Hart, Kropp, & Laws, 2004). Some are pure actuarial tools that present a fairly one-dimensional perspective on an individual's relative risk level (Hanson, 1997) while others are designed to be utilized as support systems for more comprehensive clinical determinations (Hart et al., 2004). Some rely solely on static variables (Hanson & Thornton, 1999), while others integrate dynamic predictors on a limited (Epperson et al., 1999) or exclusive (Hanson & Harris, 2000a) basis.

Following a brief review of some of these instruments, we will consider the relative utility of the various approaches.

Actuarial Assessment

Several specialized actuarial instruments for the prediction of sexual re-offense have emerged during the past decade. The actuarial approach, in a nutshell, gathers a series of variables believed to have predictive validity, applies relative weights to each variable, and combines these data into an aggregated risk score and classification.

One widely used instrument is the Rapid Risk Assessment for Sexual Offense Recidivism, known as the RRASOR (Hanson, 1997). The RRASOR is notable for its brevity and ease of use—it consists of only four variables, all of which can be easily pulled from official records. These four factors—prior sexual offenses, extra-familial victims, offender age under 25, and male child victims—were identified for use in the scale based on research indicating a strong correlation between these factors and risk of re-offense. While demonstrating moderate predictive accuracy, the RRASOR omits several variables shown to have particularly high correlations with re-offense risk, including deviant sexual preferences, antisocial orientation, and treatment compliance.

A second commonly used tool, the Static-99, addresses some of these shortcomings by combining the RRASOR with a second scale, the Structured Anchored Clinical Judgement-Minimum. Beyond the variables contained in the RRASOR, the Static-99 considers a range of additional factors including sexual deviance, range of available victims, persistence, and a pattern

of antisocial behaviors (Hanson & Thornton, 1999; Hanson & Thornton, 2000). In a comparative review, the Static-99 has been demonstrated to add to the predictive accuracy of the RRASOR in the measurement of long-term risk potential (Hanson & Thornton, 2000).

A third instrument—the Sex Offender Risk Appraisal Guide (SORAG) (Quinsey et al., 1998)—measures a different, although likely closely related group of factors compared to the Static-99. This scale, adapted from a general violence prediction tool known as the VRAG, is notable for its integration of psychiatric and psychological variables, including psychopathy and mental illness diagnoses. Its relative predictive value appears comparable to the Static-99 in the prediction of sexual recidivism, and appears to more effectively predict non-sexual violent recidivism (Hanson & Thornton, 2000).

Beyond these instruments designed for general use, some states have developed customized instruments, generally under the auspices of a state agency, designed for specific uses. The Minnesota Sex Offender Screening Tool (MnSOST) was originally developed in the early 1990s by the Minnesota Department of Corrections as a means of codifying factors viewed to place an individual at high risk for re-offense (Huot, 1999). Revised to the MnSOST-R in 1996 (Epperson et al., 1999), the tool was explicitly designed to be used by non-clinical staff. Research on the MnSOST has demonstrated moderate predictive capacity, comparable to other commonly used actuarial instruments (Barbaree, Seto, Langton, & Peacock, 2001; Hanson & Morton-Bourgon, 2004).

Finally, the SONAR (Sex Offender Needs Assessment Rating) was designed in 2001 as an actuarial tool based on dynamic variables (Hanson & Harris, 2000a). Viewed as an adjunct to actuarial instruments based on static factors, the SONAR captures information across both stable and acute dimensions. Stable factors include intimacy deficits, negative social influences, attitudes toward sex offending, and self-regulation. Acute factors include substance abuse, negative moods, anger, and victim access. The SONAR was subsequently adapted into two scales—the STABLE 2000 and the ACUTE 2000. These scales, combined with the Static-99, form the basis for a blended approach toward community supervision designed to capture long-term, intermediate, and short-term factors associated with sexual recidivism (Harris & Hanson, 2003).

Structured Clinical Decision Tools

In contrast with actuarial instruments, which contain explicit rules for weighting each variable, structured clinical assessment guides the evaluator to consider a range of empirically validated risk factors, which the evaluator then assesses for a general estimate of risk.

One example of a structured clinical decision tool is the Sexual Violence Rating Scale (SVR-20) (Boer, Hart, Kropp, & Webster, 1997). Applying a similar approach to the HCR-20—a tool used to structure clinical decisions regarding the risk of general violence (Webster et al., 1997)—the SVR-20 encompasses twenty variables that are distributed into three broad domains. These domains include psychosocial adjustment (encompassing factors such as sexual deviance, history of childhood sexual abuse, psychopathy, relationship problems, employment instability, and offending history); the nature of sexual offending (such as levels of violence employed, escalation in offense severity, and attitudes toward offending behaviors); and future plans (i.e., responses to interventions). One notable characteristic of tools such as the SVR-20 lies in their potential to capture and integrate the individual's responses and reactions to treatments and interventions. The SVR-20 has recently been modified into a new instrument known as the Risk for Sexual Violence Protocol (RSVP) (Hart et al., 2004).

Comparing the Approaches

The research on sexual offense recidivism has focused primarily on issues of long-term risk. Bounded by certain methodological limitations, this research has been highly focused on static factors that have been demonstrated to be associated with the probability of future sexual offenses.

In the context of these circumstances, it is not terribly surprising that actuarial assessment has carried the day. Research on the comparative ability of these approaches to predict general recidivism in a population of sexual offenders has found actuarial assessments to be most accurate, followed by guided clinical approaches, then by unstructured clinical judgment (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2004).

Regarding adjusted actuarial methods, actuarial “purists” argue that the track record of clinical judgment is so poor that scores on validated instruments should not be tainted with any subjective interpretation (Quinsey et al., 1998). However, some have suggested that the proven superiority of such adjusted approaches in other domains (notably weather forecasting) suggests that the adjusted actuarial approach may represent a promising approach (Monahan & Steadman, 1996; Swets, Dawes, & Monahan, 2000). To date, however, little or no empirical evidence has emerged testing this premise (Hanson & Morton-Bourgon, 2004).

Yet before concluding that actuarial methods represent a uniformly superior means of risk assessment, one must also recognize certain limitations to the actuarial approach.

First, on a purely conceptual level, many have questioned the validity of making case-level determinations on the basis of tools that have been derived solely from population-based probabilities. This issue has arisen particularly in the context of sexually violent predator civil commitment decisions. Accordingly, some have argued that sole reliance on actuarial instruments risks missing important clinical information that can aid significantly in prediction (Hart, 1998).

Second, the most widely employed actuarial scales are implicitly one-dimensional in nature. Conceivably, individuals with entirely different constellations of risk factors may be classified at similar levels of risk—a fact that may obscure important distinctions for purposes of service planning and supervision. In this sense, actuarial scales’ reliance on measures of cumulative risk obscure important case characteristics that may indicate an elevated risk. While some have called for multi-dimensional models drawing from multiple actuarial instruments (Doren, 2004b), evidence to date has not lent support to such an approach (Seto, 2005).

Ultimately, the relative superiority of one method or another is highly dependent on the questions that we are asking. If our primary concern deals with the aggregated long-term risk posed by a group of individuals, actuarial instruments almost certainly provide the most valid means of assessing such risk. If we are concerned with setting forth the relative *probability* that a particular individual will re-offend at some undetermined point in the future, actuarial instruments provide a moderate degree of accuracy, albeit one prone to errors.

Yet as soon as we turn to different types of questions, the relative utility of currently available actuarial instruments dissipates considerably. Under what circumstances would this person be most likely to reoffend? What is the probable timeframe of re-offense? How has this person’s re-offense risk been mitigated by our interventions? What is the probable impact of treatment and supervision? While work continues on actuarial approaches that might eventually answer some of these questions, these issues simply cannot be adequately addressed by current actuarial methods.

Considering these factors, the remainder of this article is grounded on three fundamental premises regarding the clinical-actuarial distinction—first, that any discussion regarding the relative merits of clinical vs. actuarial approaches cannot occur in a vacuum, and must be placed in its appropriate programmatic and operational context; second, that, while the clinical-actuarial distinction is important from a theoretical perspective, and while some circumstances call for orthodox adherence to one of the two methods, the majority of sex offender management practice calls for operating on a “middle ground” that draws from both approaches; and third, that the clinical-actuarial continuum is only one dimension within a broader practical framework that integrates a range of related constructs.

[back to top](#)

Risk Assessment and Community Supervision Practice: A Framework

Having reviewed the existing state of sexual offender risk assessment knowledge and practice, we now turn to the fundamental question presented at the outset of this article—how can risk assessment systems and methods be effectively aligned with the specific goals and challenges of community supervision practice?

As noted earlier, the risk assessment methods to be applied in a given situation are highly dependent on the specific questions that need to be answered. The variability of these questions may be viewed in terms of *organizational demands*, in accordance with the distinct information needs of central management, unit supervision, and line staff; in terms of *case level demands*, recognizing the significant heterogeneity of the sex offender population in terms of offense type and severity, motivations, and associated levels of risk; and in terms of *temporal demands*, noting that the challenges associated with managing a particular case change over time.

[Figure 1](#) presents a multi-dimensional framework that aims to address some of these sources of variability, integrating the risk assessment concepts described earlier with the range of challenges associated with the community supervision of sexual offenders.

Key Dimensions

The five noted dimensions, described below, are not intended to be categorical in nature—rather, each should be viewed as a part of a continuous spectrum of choices that must be made in conjunction with the process of community supervision.

Dimension #1: Primary Orientation

The model begins by framing the distinction between a nomothetic (i.e. population-based) approach and an idiographic (i.e. individual-based) approach as a framework with which to understand the appropriate (and in turn the inappropriate) application of risk assessment to sexual offenders. Under a nomothetic orientation, decisions are driven exclusively by population-based probabilities based on empirically validated systems of evaluating risk. Evidence-based policy and practice in this arena involves the assessment of the populations of concern using actuarial methods that rely primarily on static or highly stable predictors of re-offense. Under an idiographic orientation, decisions are based on case-specific attributes based on circumstances presented during a particular point in time. Evidence-based practice in this instance depends on the far less developed area of research into dynamic predictors of sexual recidivism, and generally employs informed practitioner judgment.

Dimension #2: Risk Emphasis

Heilbrun (1997) distinguishes between prediction-oriented styles and management-oriented styles of communicating risk. Prediction-oriented risk assessments are appropriately applied in contexts that explicitly call for understanding general likelihood of an event occurring at some undefined point in the future. Conversely, management-oriented approaches are more suited to the ongoing task of understanding and managing risk at the case-level. Dvoskin & Heilbrun (2001) associate the predictive orientation with actuarial methods and the management orientation with clinical methods, suggesting this distinction as a means of bridging the divide in the actuarial vs. clinical debate.

Dimension #3: Risk Factors

The framework presents risk factors as a spectrum covering three general domains—static factors, stable (dynamic) factors, and acute (dynamic) factors (Hanson & Harris, 1998). As a matter of practice, there is a strong rationale for viewing these factors as a continuum rather than as discrete categories. While some factors such as historical variables are by definition immutable, other “static” variables might straddle the domains. For example, whether psychopathy falls in the category of a static, immutable category or a highly stable (but ultimately changeable) personality characteristic remains open for debate. The precise boundaries

between stable and acute dynamic factors, often loosely defined by issues of timing or magnitude, may be similarly unclear.

Dimension #4: Primary Methods

The actuarial-clinical dimension, as the other dimensions contained within the framework, is viewed as a spectrum of alternatives, rather than an “either-or” proposition. At the far ends of this spectrum, the specified method is framed as the predominant (although not necessarily exclusive) means of gathering salient and valid information. In the middle of this spectrum, the framework considers blended approaches integrating both actuarial and bounded practitioner judgment as the most effective means of assessing risk.

Dimension #5: Frequency of Assessment

The final dimension—the frequency of risk assessment processes—represents a critical operational issue related to planning and implementation of community supervision systems. Baseline assessments based on exclusively static variables, by definition, tend not to require repeated administration, presenting minimal burden on operations and resources. Similarly, periodic structured assessments to gauge gradual change in relatively stable case characteristics can be integrated into regular work processes with predictable impact. The ongoing demands associated with identifying and responding to imminent risk, however, present a wide range of operational challenges associated with issues such as communication, surveillance systems, and staff workloads. This factor may also be viewed as a significant potential operational impediment to the introduction of actuarial methods as a means of assessing acute risk.

Policy and Practice Domains

Having briefly considered the general dimensions and the relationships between them, our next step is to apply these dimensions to the specific challenges associated with the community supervision of sexual offenders. The figure’s primary columns, denoted by the shaded boxes, divide these challenges into four general domains—policy and management, baseline planning, case management, and acute intervention.

Policy and Management

The policy and management domain encompasses the actions and decisions of organizational leadership within parole and probation agencies. It may also, under certain circumstances, encompass the actions of legislators charged with the crafting of public policies associated with community-based sex offender management.

Although actors within this domain generally operate independently of case-level decisions, reliable and valid data regarding the risk levels presented by individuals under agency supervision emerge as vital management indicators—indicators that affect such matters as the formulation of policies and procedures, the allocation of resources, organizational strategy, quality management, and program design.

At the policy and management level, effective decision-making depends largely on the maintenance of a nomothetic perspective centered upon population-based indices and patterns. The information required to make key decisions in this domain emphasizes the prediction of general risk within the population, rather than the specific risk posed by individual cases. Consistent with these goals, actuarial assessments driven by static characteristics within the population generally provide appropriate levels of information to inform decisions regarding resources and organizational strategy.

Baseline Planning

At the individual case level, one of the initial tasks faced by community supervision agencies involves the establishment of baseline levels of risk. This assessment may occur as part of a Pre-Sentence Investigation (PSI) process, as part of a prisoner re-entry plan, or as part of the

agency's classification process, often in collaboration with law enforcement, correctional authorities, and treatment providers.

Within this domain, line managers and staff are required to make a range of decisions associated both with the initial terms of probation or parole and with the allocation of often limited resources. Who is appropriate for lifetime or intensive supervision? What special conditions and restrictions need to be placed on each individual? What are each individual's treatment needs and potential responsiveness to treatment? Whose risk might be mitigated with access to ancillary services such as substance abuse treatment, employment programming, mental health services, or residential programming?

Within this domain, prediction-oriented assessments provide case managers with a baseline assessment of an individual's general risk. Such general predictions may inform such decisions as development of initial case plans and the resource-intensity of supervision. Over time, however, prediction-oriented notions of risk gradually lose their relative utility to those charged with individual supervision, giving way to a significant demand for management-oriented approaches.

Case Management

In contrast with the baseline planning stage, the case management domain shifts the emphasis from the realm of prediction into the realm of management. While baseline risk levels provide highly relevant context to ongoing service planning and risk management, the greater concern becomes the flow of information regarding changes in the offender's psychological, social, or environmental status. Are insights and attitudinal adjustments being gained in treatment? Has the individual managed to maintain relationships, employment, and housing? Does the process of community integration seem to be succeeding?

The answers to these types of questions carry a range of implications for both the agency and for the individual case. At the agency level, they help to prioritize the assignment and allocation of resources, and provide potentially valuable information to managers regarding the efficacy of interventions. At the case level, they provide supervision staff with vital data relevant to adjustments in service plans, expansion or contraction of terms and conditions, or identification of emergent needs.

Acute Intervention

In contrast with the case management domain, in which programmatic adjustments are made based on gradual evolution of circumstances, decisions in this domain are concerned with short-term changes in psychological, social, or environmental conditions that might presage offending behavior. This domain's primary concern is based on one central question—namely, when is an individual at imminent risk of re-offending?

By necessity, this domain focuses on the unique characteristics of the individual case, and accordingly falls at the idiographic end of the spectrum. While knowledge of the individual's general risk level might provide useful context, general predictions of the person's probability of re-offense are far less salient than information that will identify factors associated with pending re-offense and, in turn, inform appropriate intervention. Accordingly, the static risk factors that might have contributed to this individual's baseline risk assessment carry relatively little practical utility when compared to time-specific situational factors such as access to potential victims (i.e., opportunity), relapse into drug or alcohol use, lapses in compliance with terms of supervision, and stressors such as the loss of a job, home, or relationship.

In contrast with domains towards the other end of the spectrum, assessment methods within this domain remain highly dependent on bounded professional judgment. While actuarial systems for evaluating changes in acute dynamic factors among sexual offenders remain under development (Hanson & Harris, 2001), their efficacy and utility have not been fully explored. Accordingly, given the current state of knowledge, the assessment of acute dynamic factors remains largely

dependent on practitioner judgment supported by effective training, protocols, and systems of communication.

Addressing Population Variation

The framework presented above suggests that risk assessment methods must adapt to variation across organizational processes and functions. Equally important, however, these methods must respond to another critical source of variation, specifically that related to the population under supervision.

The heterogeneity of the sex offender population is well established, and has been delineated in a range of typologies developed in both the clinical and law enforcement context. These typologies have identified significant areas of divergence within populations of rapists (Knight, 1999), child molesters (Knight, Carter, & Prentky, 1989; Lanning, 1986) and even sexual murderers (Schlesinger, 2004). Key dimensions associated with this variation include primary motivation, intelligence, underlying sexual deviance, anger, opportunity, victim relationship, level of force, and a broad range of other factors.

Within sex offender typologies, the various constellations of these factors (some of which may be closely related) produce a wide range of potential offender subtypes, each associated with distinctive levels of risk. Further, and perhaps more critically, the triggers for re-offense may be markedly different across these subtypes. This factor presents significant challenges to the development of uniform methodologies for assessing dynamic risk.

Of particular importance to community supervision practice is the fact that, to date, the evidence base has been weighted towards higher-risk offenders, particularly those who have been released following a prison sentence. Accordingly, the dynamic factors that may trigger re-offense among probationers who fall into lower long-term risk categories are far less understood, and represent a critical area for future research (Hepburn & Griffin, 2004; Meloy, 2005).

[back to top](#)

Conclusion

Approximately 60 percent of sex offenders under correctional supervision in the United States are sent to serve their sentences in the community (Greenfeld, 1997). Moreover, sexual offenders comprise approximately 5-6 percent of individuals released to parole agencies, with an estimated 30,000 under parole supervision on a given day (Hughes, Wilson, & Beck, 2001). Considering these figures, the development of effective systems for community-based supervision looms large in our overall approach to sex offender management.

As more and more jurisdictions develop specialized capacity to manage sex offenders in the community (English et al., 1997), the demands for effective risk assessment have continued to expand. As such, it remains vital that the role of risk assessment, and consequently the methods that are employed, be bounded by the *specific challenges* faced by probation, parole, and community corrections agencies. This requires recognizing and adapting to the range of variation both in organizational-programmatic goals and within the population under supervision.

On a final note, the role of risk assessment in community supervision practice cannot be divorced from the unique social and political context in which our society views sexual crime, its perpetrators, and its victims. With the issue of sexual offending remaining at the forefront of legislative agendas, and with persistent public misconceptions surrounding the nature of sexual crime, community supervision agencies operate in a political environment with a “zero tolerance” approach to errors, in which one tragic case can lead to widespread calls for system reform. In this environment, the imperative of targeted, adaptable, and responsive means of risk assessment should be evident.

[back to top](#)

[References](#)

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

Figure 1: Framework for Risk Assessment in Community Supervision

Service Domains	Policy & Management <ul style="list-style-type: none"> • Strategy • Resource Allocation • Program Design • Quality management Baseline Case Planning <ul style="list-style-type: none"> • Classification • Special Conditions • Terms of Supervision 	Ongoing Case Management <ul style="list-style-type: none"> • Sex Offender Treatment • Housing & employment • Ancillary Services • Case Plan Adjustments 	Acute Interventions <ul style="list-style-type: none"> • Enhanced Supervision & Surveillance • Crisis Intervention • Police Action
Primary Orientation	Nomothetic		Idiographic
Risk Emphasis	Prediction of General Risk		Management of Specific Risk
Applicable Risk Factors	Static	Stable	Acute
Primary Methods	Actuarial	Practitioner	Judgment
Frequency of Assessment	Once at Baseline	Periodic	Ongoing

The framework consists of two sets of elements—a series of key dimensions, denoted by the five black bands, and a series of domains, contained within the grey shaded box.

[Home](#)

Validating the Level of Service Inventory—Revised on a Sample of Federal Probationers

[References](#)

Anthony W. Flores, M.S.

California State University, Bakersfield

Christopher T. Lowenkamp, Ph.D.

Paula Smith, Ph.D.

Edward J. Latessa, Ph.D.

University of Cincinnati

[Method](#)

[Results](#)

[Discussion](#)

THE LAST TWO DECADES have borne witness to a rise in the correctional population so colossal that it was previously inconceivable to practitioners and criminologists alike. As a result, annual exercises in economic calisthenics have become common practice for corrections administrators throughout the U.S. The complexity of such a population boom can only be fully appreciated by realizing that available resources have hardly kept pace. Expectations to do more with less, an emerging mantra among correctional administrators and practitioners, drive corrections professionals in a seemingly unending search for promising technological developments that might help to bridge the service gap created by surging offender populations and waning budgets.

The deleterious effects of this current “crisis in corrections” are becoming so entrenched in local- and state-level practices that some have remarked, “Corrections has become the Pac-Man of government budgets, gobbling up resources as legislators seek to finance competing needs with shrinking tax revenues” (Pierce, 1991 as cited in Cullen, Wright, & Applegate, 1996, p. 70). A discussion of the burdens levied by this crisis would be remiss if it overlooked the presence of these issues at the federal level. In fact, according to a most recent year-in-review report, the federal probation and pretrial services system has recently suffered severe budgetary and, consequently, operational constraints (United States Probation and Pretrial Services, 2005). Specifically identified were operational and service restrictions resulting from unprecedented budget deficiencies. In an attempt to realize some relief through cost-saving practices, the internal administration of federal probation and pretrial services recommended that: a) the agency move toward an evidence-based approach and implement a process to measure client outcome that will generate meaningful agency evaluations and subsequent operation refinements; b) the agency strive to restructure staff workloads and prioritize resources, most notably by identifying offenders eligible for early release; and c) the agency remain resolute in its commitment to provide services of exceptional quality in spite of fewer resources (United States Probation and Pretrial Services).

Actuarial classification systems that yield valid measures of risk and criminogenic need hold

considerable promise for correctional agencies in these regards. First, beyond informing decisions about custody and service provision, initial and reassessment risk/need scores also provide outcome measures useful for evaluating both offender and agency success. Second, because classification tools identify different levels of offender risk, the tools' corresponding risk categories are inherently useful for restructuring staff workload, prioritizing agency resource expenditure, and identifying low-risk offenders for early release. Finally, actuarial classification systems can do a great deal for agencies concerned about the quality of service provision. Classification tools can improve service quality by promoting resource, custody, supervision, and treatment decisions that are better informed, more accurate, and ultimately more useful (Latessa, 2003–2004).

Classification

Classification systems disaggregate heterogeneous correctional populations into subgroups that maximize between group differences and minimize within group differences. Classification processes create these subgroups on the basis of offender characteristics relevant to correctional outcome, which in turn facilitates and justifies differential service provision. The use of classification devices allows correctional agencies to simultaneously address multiple objectives, including improved predictive accuracy, better informed treatment assignments, more effective supervision, and meaningful outcome analysis (Clements, 1996). While most correctional agencies currently use some type of classification system to guide decision making (Jones, Johnson, Latessa, & Travis, 1999), it is important to note that all classification systems are not created equal.

The approach taken to behavioral prediction can significantly affect the resultant validity. Decisions involving behavioral forecasts are grounded in either a clinical or actuarial approach (Meehl, 1954). Whereas clinical decision making is intuitively based and justified by claims of training, experience, and expertise, actuarial decision making is based on scientific evidence generated from observed behavioral outcomes for risk-similar offenders (Monahan, 1981). Research investigating the accuracy of each method consistently supports the superiority of actuarial classification decisions over clinical practices (Gottfredson & Gottfredson, 1986; Grove, Zald, Lebow, Snitz, & Nelson, 2000).

In addition to the importance of *how* predictions are reached, evidence also suggests that *what* is assessed can greatly affect the accuracy and utility of classification decisions. While the majority of actuarial classification devices rely exclusively on historically-based static risk factors (e.g., criminal history) to predict reoffending, others have evolved in response to recent knowledge advancements in risk prediction. Heeding the findings evidenced in current risk factors research (for example, see Gendreau, Little, & Goggin, 1996), researchers have developed advanced classification tools to augment traditional static-driven prediction models with a battery of predictors said to be dynamic because of their present-day (as opposed to life-history) focus. The resultant combination of both static and dynamic risk factors yields greater predictive accuracy for classification systems. For example, in an examination of male parolee revocation, Brown (2002) found that combined static and dynamic prediction models were significantly better at forecasting revocation than either static or dynamic models alone. Additionally, in a quantitative literature review comparing the averaged correlation between static risk factors and recidivism to the averaged correlation between dynamic risk factors and recidivism, findings supported the greater predictive accuracy of dynamic factors (Gendreau et al.).

Beyond improving the accuracy of predictions, devices that measure risk through a combination of static and dynamic predictors offer considerable utility to correctional agencies. Because dynamic risk factors are characteristics currently present in an offender's life, they are inherently sensitive to detecting change and measuring progress (Taylor, 2001). Measuring both static and dynamic risk factors thus becomes a way of not only improving risk management but moving toward risk reduction (Latessa, 2003-2004). That being said, it is important to note that risk reduction depends upon correctional agencies matching their service intensity levels according to actuarial-determined offender risk levels (Lowenkamp & Latessa, 2005). Empirical investigations have consistently established that to be effective, correctional agencies must direct a majority of

their resources toward high-risk offenders. Correctional practices informed by this approach are commonly referred to as adhering to the risk principle; they offer evidence-based services entrenched in research findings that have consistently demonstrated efficacy in reducing offending while also avoiding the iatrogenic effects associated with servicing low-risk populations (Andrews, Zinger et al., 1990; Dowden & Andrews, 1999a, 1999b; Lowenkamp & Latessa, 2004a; Lowenkamp & Latessa, 2004b).

The Level of Service Inventory-Revised

One example of an actuarial classification system that measures risk and criminogenic need is the Level of Service Inventory-Revised (LSI-R). The LSI-R measures 54 risk and need factors about 10 criminogenic domains that are designed to inform correctional decisions of custody, supervision, and service provision (Andrews & Bonta, 1995). The theoretically informed predictor domains measured by the LSI-R include criminal history, education/employment, financial situation, family/marital relationships, accommodation, leisure and recreation, companions, alcohol or drug use, emotional/mental health, and attitudes and orientations (Andrews & Bonta).

The LSI-R assessment is administered through a structured interview between the assessor and offender, with the recommendation that supporting documentation be collected from family members, employers, case files, drug tests, and other relevant sources as needed (Andrews & Bonta, 1995). The total risk/need score produced by the LSI-R is indicative of the number of predictor items (out of 54) scored as currently present for the offender. The LSI-R score is then actuarially associated with a likelihood of recidivism that was derived from the observed recidivism rates of previously assessed offenders. Last, domain scores of the LSI-R are used to identify an offender's most promising treatment targets (Andrews & Bonta).

Because the LSI-R represents a theoretically informed, empirically supported, actuarial-based, and standardized measure of criminogenic risk and need, it boasts considerable potential to improve caseload decisions, resource expenditure, and overall service quality (Andrews & Bonta, 2003; Gendreau et al., 1996).

The LSI-R and Validity

The benefits promised by any classification system must be empirically evaluated against the benefits actually observed in prior research evaluations. Research findings have generated a significant body of evidence that established the LSI-R as a valid predictor of correctional outcome across a variety of measures. Specifically, findings have supported the predictive validity of the LSI-R for institutional infractions (Bonta, 1989), probation failure (Andrews, Kiessling, Robinson, & Mickus, 1986), halfway house failure (Bonta & Motiuk, 1985; Motiuk, Bonta & Andrews, 1986), and parole violations (Bonta & Motiuk, 1990). Current validity research on the LSI-R has also supported the tool as a promising predictor of future offending (Andrews & Bonta, 1995; Goggin, Gendreau, & Gray, 1998). Moreover, empirical analyses reveal that the instrument's accuracy in predicting future offending holds across correctional settings and offender populations (Holsinger, Lowenkamp, & Latessa, 2004; Lowenkamp, Holsinger, & Latessa, 2001).

There are, however, warranted concerns about the population-specific nature of prediction tools. For instance, Wright, Clear, & Dickson (1986) tested the predictive validity of the Wisconsin model risk assessment on samples of probationers in New York and Ohio. Though the Wisconsin model had demonstrated predictive validity for the sample upon which it was created (Baird, Hines, & Bemus, 1979), it failed to demonstrate predictive validity for either the New York or Ohio sample of probationers (Wright et al.). Specific to the LSI-R, research conducted by Dowdy, Lacy, & Unnithan (2001) failed to support the tool as a predictor of halfway house outcome, two-year recidivism for any crime, or two-year felony recidivism for a sample of halfway house offenders. Taken together, these two findings serve as a reminder to correctional agencies that classification systems must be validated to their specific offender populations.

The literature provided on the validity of the LSI-R has established the tool as a valid predictor of correctional outcome across offender types and settings. The information obtained from the LSI-R can increase the accuracy of important corrections-related decisions (i.e., classification, risk level, criminogenic needs, service provision, intensity of interventions, and program effectiveness). In a similar vein, assessment research has also indicated a lack of universal applicability for prediction instruments. Federal probationers represent a unique correctional population in that they are older, more likely to be Hispanic, and more likely to be drug offenders than their state-supervised counterparts (Glaze & Palla, 2005; United States Probation and Pretrial Services, 2005). Because of this, and because of a current lack of existing research on the validity of the LSI-R for federal probationers, this research investigated the predictive accuracy of the LSI-R for a sample of federal probationers.

[back to top](#)

Method

Participants

The sample in this study was comprised of 2,107 adult federal probationers. To be eligible for inclusion, a federal probationer had to have been assessed with the LSI-R by a federal probation staff member trained in the administration and scoring of the tool. LSI-R assessment scores for the sample were completed over a two-year period between December of 2001 and December of 2003.

Procedures

In 2001, the southwestern federal probation district that provided these data received a three-day training on the implementation and scoring of the LSI-R. Six months later, follow-up LSI-R training was provided for all staff and immediately followed by a “train the trainers” session for staff that had demonstrated exceptional LSI-R scoring skills. During the follow-up and “train trainers” sessions, administrative staff voiced concern about the tool’s applicability to federal probationers and expressed interest in norming and validating the LSI-R on their offender population. This early discourse between federal probation staff and research consultants about the LSI-R’s psychometric properties served as the impetus for what later matured into a collaborative effort between both parties to provide the agency with aggregated probationer needs reports, normative information for their offender population, and evidence attesting to the LSI-R’s validity for federal probationers.

Save for outcome, the variables of interest in this study were entered into an electronic database maintained by federal probation staff. Once the number of offenders in the database exceeded 2,000, federal probation staff sent a copy to the authors (via electronic mail). Upon receipt, the data were cleaned and then used to generate a data collection sheet that individually listed all sample participants by their name, age, sex, ethnicity, race, and county of committing offense. These data collection sheets were then used to collect outcome data for the sample.

Measures

Although the LSI-R is comprised of ten risk and criminogenic need areas, only the composite LSI-R score was used in the current research. The LSI-R scores used are the result of offender interviews and collateral reviews of file and other offender information as completed by a federal probation staff member. Recidivism data were collected by completing follow-up record checks for each offender in the Federal Bureau of Prisons’ inmate locator database. The measure of recidivism used in the current study was incarceration in the Federal Bureau of Prisons for either a technical violation or new offense that occurred subsequent to the initial LSI-R assessment. Recidivism was coded dichotomously, where a value of 1 indicated the occurrence of subsequent incarceration and a value of 0 indicated that subsequent incarceration had not occurred.

Several demographic variables were also used in the current analyses to both describe the sample of offenders studied and further specify the relationship between the LSI-R and incarceration

through their consideration in a multivariate analysis. The additional variables included in the multivariate analysis were age, sex, and ethnicity. While age was a continuous variable, sex and ethnicity were both coded dichotomously, so that a value of 0 represented the most typical case (male and Hispanic, respectively) and a value of 1 represented a departure from the most typical case (female and non-Hispanic, respectively).

It should be noted that the Federal Bureau of Prisons' database used to collect recidivism data only allowed for the determination of incarceration in a federal institution subsequent to the initial LSI-R assessment. Alternate measures of outcome, such as commission of a technical violation, time until new commitment, or incarceration under state jurisdiction were unavailable. Though the use of additional outcome measures would admittedly yield more information about the length of time until and type of recidivism, the use of subsequent incarceration is advantageous to the extent that it provides a more conservative test of the LSI-R's predictive validity.

[back to top](#)

Results

Descriptives

Descriptive statistics for offender demographics, LSI-R scores and incarceration are presented in [Table 1](#). An examination of these data reveals that the typical federal probationer in this sample was a Hispanic male, of 37 years, classified as low/moderate risk on the LSI-R ($M = 14.08$, $SD = 7.81$). The descriptive results also revealed a 26.1 percent base rate of incarceration for the sample, indicating that nearly three out of four federal probationers had not recidivated prior to the completion of follow-up record checks for the sample.

Validation

The predictive validity of the LSI-R for federal probationers was examined by conducting three separate analyses. The first of these involved calculating a predictive validity estimate of the relationship between the LSIR and incarceration for the sample of 2,107 federal probationers. This analysis supported the LSI-R as a significant predictor of subsequent incarceration ($r = .283$, $p < .01$).

The second test in this research investigated the validity of the LSI-R and incarceration using receiver operating characteristic (ROC) analysis. The correlation coefficient calculated in the first analysis of this research represents the validity estimate most commonly reported in existing LSI-R research endeavors. However, because the magnitude of a correlation coefficient is dependent on the percentage of the sample identified as a recidivist by the risk tool (selection ratio) and percentage of actual recidivists in that sample (base rate), it can be said that existing research has yet to yield a statistic that would permit an unbiased comparison of the LSI-R's predictive strength across samples and as compared to other prediction tools. To address this deficiency, ROC analysis was chosen as a second means of examining the LSI-R's predictive validity in this study. The ROC method produces a statistic that, because it is unaffected by sample-specific base rates and selection ratios, is remarkably useful in comparing the utility and strength of a prediction instrument across different samples and vis-à-vis other prediction tools (Mossman, 1994a).

Statistics derived from ROC curves represent the ratio of true positives to false positives present in a given sample (i.e., as identified from predictions made by the risk instrument and observations for the outcome variable). The ROC analysis completed in this research produced an area equal to .689 ($p < .01$) to describe the relationship between the LSI-R and subsequent incarceration. Simply put, there was a 68.9 percent chance that a randomly selected recidivist had a higher score on the LSI-R than did a randomly selected nonrecidivist (Rice & Harris, 1995).

The analyses performed thus far have revealed that the LSI-R was a valid and robust predictor of

subsequent incarceration for federal probationers. To further specify this relationship, the final analysis in this study estimated a multivariate logistic regression model that examined the relationship between the LSI-R and incarceration while simultaneously considering the effects of age, sex, and ethnicity. The results of this multivariate analysis are reported in [Table 2](#). An examination of the logistic regression model reveals that the LSI-R continued to be a significant predictor of incarceration, even when offender age, sex, and ethnicity were controlled. Moreover, a review of the values for *Exponent(B)* shows that the LSI-R was the strongest predictor of incarceration among the variables included in the multivariate model. This conclusion is also reached through an examination of the *R* values estimated for each of the predictor variables. The *R* value of .250 for the LSI-R is more than twice the *R* values estimated for the other significant predictors of incarceration included in the model (age and sex). Finally, the nature of the relationship observed between incarceration and the control variables of age and sex is also worth noting. The logistic regression analyses revealed a significant and negative *R* value for each variable, indicating female federal probationers and older federal probationers were less likely to recidivate than were their male and younger counterparts.

[back to top](#)

Discussion

This research sought to make two important contributions to the knowledge base of recidivism prediction. First, because the utilization of valid and efficient risk/needs assessment tools implies a certain level of improvement in the ability to manage offender caseloads and classify correctional populations, their use is becoming increasingly prevalent throughout North America. As the popularity of such prediction tools increases, so too, does the diversity of the offender population to which the tools will be applied. Consequently, it was deemed important to examine the LSI-R's efficacy to predict long-term outcome for a sample of federal probationers, a correctional population previously overlooked in existing LSI-R validation studies. Second, this study used ROC methods to calculate an unbiased measure of predictive accuracy appropriate for comparisons across samples and between different tools, a measure absent in existing research, but critical for knowledge advancement.

Results from the predictive validity analyses were encouraging and provided evidence that the LSI-R was a valid and robust predictor of subsequent incarceration for this sample of federal probationers. Additionally, the multivariate analysis conducted in this research found that the LSI-R remained a valid predictor of subsequent incarceration when the effects of age, sex, and ethnicity were controlled. Taken together, these results make a strong case for the generalizability of the LSI-R to diverse offender populations. The findings of these, as well as previous, analyses further demonstrated that a theoretically informed and empirically refined actuarial measure of client risk and criminogenic need has much to offer correctional practice. When these findings are considered in the context of existing research, use of the LSI-R to inform correctional decisions of supervision and service provision appears to epitomize evidence-based practices.

In addition to further establishing the predictive validity of the LSI-R, these analyses contributed to existing research by using ROC methods to calculate an index of predictive accuracy that is independent of sample base rates and selection ratios. ROC methods are important in prediction research to facilitate comparisons of predictive strength across samples and, more importantly, across prediction instruments. The ROC area of .689 generated in this research was moderate to large in magnitude (Rice & Harris, 1995), and indicated that a randomly selected recidivist would have a higher LSI-R score than would a randomly selected nonrecidivist 68.9 percent of the time. Beyond the importance of this finding for the accuracy of the LSI-R, it is hoped that the ROC analyses reported here prompt future prediction efforts (on all risk/need assessment tools) to consider the inherent link between comparisons of predictive accuracy and knowledge advancement.

Optimistic conclusions aside, there were several limitations present in this research. First, the outcome measure used did not permit analyses with respect to length of time until failure, failure

due to technical violation, type of crime committed at failure, or failure under state jurisdiction. Additionally, these analyses examined the predictive accuracy of the LSI-R for a large sample of federal probationers, but did not investigate the tool's applicability to offender subgroups. Certainly, existing research can benefit from future efforts that might further specify the relationship between the LSI-R and outcome across offender sex, ethnicity, and race.

[back to top](#)

[References](#)

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

Table 1: Descriptives for the Sample (N = 2,107)

Variable		M	SD
Age		37.27	11.47
LSI-r Score		14.08	7.81
		N	%
Sex			
	Male	1,551	73.6
	Female	556	26.4
Ethnicity			
	Hispanic	1,369	65.0
	non-Hispanic	738	35.0
Recidivism			
	No Reincarceration	1,557	73.9
	Rincarceration	550	26.1

Table 2: Logistic Regression Model Predicting Incarceration for the Sample (N = 2,107)

								95% CI for Exp(B)	
Variable	B	SE	Wald	DF	Sig	R	Exp(B)	Lower	Upper
Age	-0.030	.005	34.8265	1	.000	-.117	.970	.961	.980
Sex	-0.815	.134	37.0412	1	.000	-.121	.443	.341	.575
Ethnicity	-0.036	.112	.1040	1	.747	.000	.964	.774	1.202
LSI-R	0.085	.007	152.2118	1	.000	.250	1.088	1.074	1.103
Consultant	-1.034	.211	23.9882	1	.000				

Note: -2 log likelihood = 2175.887; $\chi^2(4) = 236.335$; $p < .001$; pseudo R² = .156.

[Home](#)

Validation of the Risk and Resiliency Assessment Tool for Juveniles in the Los Angeles County probation System

[References](#) | [Endnotes](#)

Susan Turner
University of California, Irvine
Terry Fain
RAND Corporation

[Background](#)[Methods](#)[Results](#)[Discussion and Conclusions](#)

RISK AND NEEDS assessment has been central to correctional operations for decades. Assessment not only helps predict offender future behavior, it can also help organizations allocate staff workload and resources. Before the late 1970s, judgments about offender risk were often subjective, based on experience or the intuition of correctional practitioners (Solomon & Camp, 1993). Objective systems began to appear in the 1970s and offered the promise of more efficient and systematic means of classification for offender risk and management than clinical intuition alone. The National Institute of Correction's model Risk Classification initiative, undertaken in the early 1980s, introduced many jurisdictions to objective case classification (Jones, Johnson, Latessa, & Travis, 1999). Today, risk and classification tools are used in a myriad of criminal justice decisions—from pretrial release to parole supervision for both juvenile and adult populations. More recent “third generation” instruments include criminogenic needs of the offender that should be addressed in order to reduce recidivism (Bonta, 1996).

One of the most critical issues for assessment instruments is their predictive validity. An instrument should be able to accurately predict which offenders will and will not recidivate. Whether an instrument is selected from a number of commercially available products (such as the Level of Service Inventory and Correctional Offender Management Profiling for Alternative Sanctions) or developed by a jurisdiction, it should be validated on the local population. The current article discusses the validation of the San Diego Risk and Resiliency Checkup on a sample of juvenile offenders in Los Angeles County.

[back to top](#)

Background

Although the Los Angeles County Probation Department routinely gathered background information on youths entering its juvenile system, no validated risk assessment was being used through the early 2000s. As part of a court settlement regarding services provided to minority youth in the county, the department was required to allocate resources for the administration of a

validated risk and needs instrument to its juvenile probationers. Of particular importance was that the instrument work well for youths of all ethnicities.

Working with a committee representing the parties of the court settlement, researchers assisted in identifying and eventually validating a risk assessment instrument to be used in the county. After surveying instruments currently in use in the United States, we determined that items used in risk and needs instruments generally fell into one of nine conceptual categories: prior and current offenses/dispositions, family circumstances/parenting, education, employment, peer relations, substance abuse, leisure/recreation, personality/behavior, and attitudes/orientation. However, many of the instruments that we found in use had not been validated on the populations to whom they were administered, so that we were unable to determine their effectiveness in distinguishing high-risk youths from low-risk youths.

We identified three instruments that had undergone validation: the Youth Level of Service Inventory (YLSI) (Multi-Health Systems Inc., 1998), the San Diego Risk and Resiliency Checkup (SDRRC) (Little, n.d.), and the Washington Association of Juvenile Court Administrators Risk Assessment (WSJCA-RA) (Washington State Institute for Public Policy, 2004). Each includes multiple items for the conceptual categories we identified, and each offered advantages and disadvantages when compared to the others. The Department favored the SDRRC, primarily because it could be administered during the intake process. It also preferred the SDRRC's emphasis on positive ("protective") factors, whereas most risk and needs assessment instruments primarily focus on risk factors. The remaining settlement parties agreed, and the SDRRC was selected as the instrument to be tested.

The San Diego Risk and Resiliency Checkup

The SDRRC consists of 60 items in six conceptual categories, half of which are risk factors and half protective factors. The conceptual categories are delinquency, education, family, peer relations, substance use, and individual factors. Each conceptual category includes five protective factors and five risk factors. Each item is scored as "yes," "no," or "somewhat." Scores from the risk and protective subscales are combined into a single resiliency score. The SDRRC also includes additional protective factors and additional risk factors that are not included in the resiliency score, but which may be used to tailor an individual's supervision. A copy of the SDRRC instrument is included in the Appendix.

One important difference between the SDRRC and most other risk and needs instruments is that a higher score on the SDRRC implies higher resiliency, i.e., a lower score corresponds to a higher risk of re-offending. Most risk and needs instruments, by contrast, associate high scores with high risk of recidivism. The SDRRC does not contain any preset cut-points for youth risk levels.

The one existing validation study of the SDRRC was performed by Little (n.d.). This study included 2,633 youths surveyed in San Diego between February 5, 1999, and March 28, 2001. The SDRRC was found to be effective in predicting future offenses (Little, n.d.). The total resiliency profile appeared superior to either of the total risk and total protective scales. The correlation between the total resiliency profile and occurrence of a subsequent offense was -0.146 ($p < .001$). Using a logistic regression model to predict follow-up offenses, Little also found age, gender, ethnicity, and prior criminal history, as well as resiliency score, to be significant predictors of re-offending (Little, n.d.).

[back to top](#)

Methods

Selecting the Sample

We wanted to assure adequate statistical power for detecting differences in recidivism rates between low-, moderate-, and high-risk youths, as well as differences between groups defined by race/ethnicity and gender. Because the SDRRC does not have any preset risk cut-points, the pilot

study proposed to divide the sample into approximate thirds defining low-, moderate-, and high-risk groups. The probability of detecting a difference in recidivism rates between the three risk groups depends upon the number of groups (in this case, three); the sample size of each group, and the spread of the true rates of recidivism. Because we did not know the true rates of recidivism for the different risk groups, we proposed three plausible “true” scenarios for the probability of rearrest at 6 months for low-, moderate- and high-risk youths.

The three were:

	Low	Moderate	High
Scenario 1:	11%	21%	27%
Scenario 2:	12%	25%	32%
Scenario 3:	12%	18%	32%

With these three scenarios, ¹ we determined that at least 120 to 140 youths in each risk level would need to be included in order to be able to detect differences. However, we also wanted to be able to detect differences for key subgroups: boys as well as girls; and for blacks, Hispanics, and white/other youths. Each of the subgroups of interest needed between 100 and 120 youths within low, moderate, and high-risk groups for adequate power. Therefore, we needed approximately 300 to 400 of each gender and each race/ethnic group.

Our final sample size target was 1200 youths for the study. This included 800 males and 400 females, and 400 each of whites, blacks, Hispanics. Because probation officers assess youths in both court- and non-court venues, we designated approximately 800 court cases and 400 non-court cases. ²

[Table 1](#) shows the full stratified target sample, with the size of the sample in each cell.

Four area offices were chosen for the assessment in order to provide county geographical representation. ³ These were Long Beach/Harbor (South), Pomona (East), Centinela (West), and Van Nuys (North). Each area office was to supply one-quarter of the target sample assessments.

Training

Probation officers volunteered for the assessment pilot. Originally 18 Deputy Probation Officers—14 field and 4 Camp Community Transition Program (CCTP)—were trained in the administration of the SDRRC. The three-day training consisted of an overview of the instrument; its application and practice; overview of evidence-based practice, including the overview of the six criminogenic needs and the eight guiding principles for risk/recidivism reduction; motivational interviewing techniques; and the actual administration of the tool. Training was conducted by staff from Justice System Assessment & Training, the firm that developed the SDRRC. Deputy Probation Officers (DPOs) were provided with an incentive of 30 minutes overtime payment for each assessment completed during the pilot.

Data Collection

Data were collected in three general areas: assessment scores, services received, and recidivism.

Assessment Scores. DPOs administered the assessments to youths. Information on each of the additional risk and protective factors that do not contribute to the overall resiliency score was also recorded. The assessment form also includes demographic variables (age, gender, ethnicity), information about proficiency in English, and criminal history.

Assessments were conducted from December 6, 2002, through October 30, 2003. A total of 1,165 youths were assessed by Los Angeles County probation officers. We also gathered information on whether the youth’s case proceeded to supervision or ended at investigation (no

further probation supervision).

Recidivism. Using the Probation Department’s databases, we obtained information on arrests during the 12 months after assessment for each subject. These data include both juvenile and adult arrests. Date of each arrest, charges, and disposition were recorded. We also used records from juvenile halls and juvenile camps to determine how many days a given youth was incarcerated during the 12 months after assessment.

We were unable to determine whether a given youth was rearrested during the follow-up period for 129 (11.1 percent) of the 1165 youths originally assessed for the study. Our final sample is 1,036 youths. Missing data were primarily due to incomplete disposition records, so we were unable to determine whether some youths were in custody (and therefore incapable of being rearrested). We found no significant differences on gender or age between the deleted cases and those in the final sample. Significantly more Hispanics, and significantly fewer blacks, were in the final sample than among the deleted cases ($p < .001$).

Weighting the Final Sample to Reflect Population of Probation Youths in Los Angeles

Our final sample did not exactly match the target sample presented in [Table 1](#). In particular, the final sample included somewhat more males, more Hispanics, and more court cases than we had originally targeted. By weighting our final analysis sample to represent the entire population of investigation and supervision cases for Los Angeles (as described below), we have adjusted for differences between the targeted sample and final sample, so that our analyses do accurately represent the gender and ethnic mix among all Los Angeles cases.

In order to weight the final sample, we obtained the frequency of all youth investigation and supervision cases for Los Angeles during the same time period as the pilot assessment, with information on youth gender, race/ethnicity, [4](#) and court vs. non-court case type (see [Appendix \[PDF, 3.33MB\]](#)). Within each combination of gender, race/ethnicity, and court vs. non-court case type, we defined a weight to be the ratio of youths in the probation population to the number of youths in the final sample. This allowed us to weight the data to reflect the entire population on these characteristics. [5](#) All analyses were conducted on the weighted final sample.

[Table 2](#)

[back to top](#)

Results

Mean Differences in Resiliency by Demographic Characteristics

SDRRC resiliency scores differed by gender, age, and ethnicity; some differences were large enough to be statistically significant (see [Table 3](#)). The most pronounced differences were for different ethnic groups, with pair-wise comparisons of whites, blacks, Hispanics, and “other” race all producing significant differences. “Other” youths (primarily Asians) had the highest mean resiliency scores, followed by whites, blacks, and Hispanics, respectively. There were also marginally significant differences between males and females ($t = 1.89, p < .06$) and between youths aged 15 or 16 and those aged 17 or 18 ($t = -1.93, p < .06$).

Characteristics of the SDRRC

The SDRRC comprises two subscales: the protective subscale and the risk subscale. Each of these contains subscales for delinquency, education, family, peer, substance use, and individual. [Table 4](#) shows the correlations between the overall resiliency score and the individual subscales. [Table 5](#) gives the correlations among the subscales. It is important to note that the total SDRRC scale reflects “resiliency.” Resiliency is defined as the net sum of protective and risk factors. Protective and risk factors are scored differently. The higher the protective score, the more protective factors the youth has. Risk scores have negative values; the more negative the value, the higher the risk. Thus we would expect positive correlations between the total resiliency score

and 1) total risk score, 2) total protective score, and 3) the subcomponents of both risk and protective scales. In fact, that is what we see in [Table 4](#). At the same time, however, we see fairly high correlations between individual subscale items (see [Table 5](#)), suggesting that they may be redundant. Redundancy among the subscales of the resiliency score was also reported by Little (n.d.) in her analysis of the SDRRC.

Relationship Between Resiliency and Recidivism as Measured by Subsequent Arrest

For each of the youths assessed, we determined whether the youth was arrested within the 12 months following the administration of the assessment. The major question for the validation study is whether scores obtained on the SDRRC are related to subsequent recidivism.

One of the issues for recidivism studies is whether or not subjects are “at risk” to reoffend. Individuals may be removed from the sample before they have a chance to reoffend—they may be sentenced to terms of incarceration during the entire follow-up period. In some cases, these individuals may be excluded from analyses, or they may be treated as censored observations. In order to determine how large a problem this might pose for the current study, we calculated the number of days youths were “on the street” from the point of their assessment until 12 months later. The vast majority of youths (over 90 percent) had at least 10 months of street time. For the remaining youths, analyses revealed that even those with very minimal “street time” (less than 2 months) were arrested. For this reason, we did not exclude any youths from our analyses of recidivism.

[Table 6](#) presents the recidivism results for the full sample. For this and other analyses, we divided the sample into approximate thirds and categorized the resulting groups as “low” (those with score 12 or less), [6](#) “medium” (those with scores between 13 and 33), and “high” (those with score of 34 or higher). [Table 6](#) shows that the scale does validate for the overall sample. Only 8 percent of “high” resiliency youths were arrested, compared with almost 36 percent of those with “low” resiliency.

Subgroup Analyses

[Figures 1, 2, and 3](#) present the results by age, gender, and ethnicity, respectively. Within each of the major racial groups, the resiliency score is significantly related to recidivism. Regardless of ethnicity, the higher the resiliency, the lower the likelihood of arrest for youths. The same holds true for males and females, and across all ages. The discriminatory power of the instrument appears to be greatest for the younger youths in the sample (age less than 15), most likely due to more variability in outcomes among younger juveniles.

Assessing Scale Properties and Recidivism

Prior analyses have examined the relationship between the total resiliency score and rearrest. In [Table 7](#) below, we present the relationship of individual subscales to rearrest. Recall that the more negative the risk score, the higher the risk. Thus we would expect a negative correlation between risk subscales and rearrest. All subscales correlate significantly with rearrest. The absolute correlation between the total resiliency score and rearrest is 0.27—similar to the correlation observed by the Washington State Institute for Public Policy (2004) for misdemeanor and felony recidivism for the Washington Pre-Screen Assessment inventory. Interestingly, it is higher than the correlation reported by Little (n.d.). Resiliency scores have a higher correlation than do their respective protective and risk subscales with only one exception (family protective factors).

Controls for Additional Factors Related to Recidivism

Earlier analyses have examined the univariate relationship between SDRRC score and recidivism. In the following analyses, we examine the relationship controlling for additional factors that may impact how well SDRRC predicts recidivism. These factors include age, gender, and race/ethnicity, as well as whether the case is supervision (vs. non-supervision) and court (vs. non-court).

[Table 8](#) presents the results from a logistic regression analysis of the total sample. We see that, even controlling for other factors that might be related to recidivism, SDRRC resiliency is still significantly related to rearrest. Other factors are also related to rearrest: age (not being in the youngest or oldest age group [7](#)), being male (as opposed to being female), being black (as opposed to being white), and being under probation supervision during the 12-month follow-up period. The overall measure of the model yielded a Wald chi-square value of 102.1 ($p < .0001$).

The relatively lower correlations between SDRRC items and rearrest for Hispanic and “other” youths observed in [Table 4](#) might suggest that the resiliency measure is not as strong a predictor for some groups as it is for others. In order to test this, we included interactions terms between race/ethnicity and resiliency in the model identified in [Table 8](#) above. Results, shown in [Table 9](#) below, confirm that resiliency is differentially related to recidivism for whites (compared with Hispanics), although not significantly for blacks or “other” youths.

One of the questions we want to answer is whether the provision of services influences the youth’s recidivism. We would expect those receiving services might have lower recidivism rates. In order to evaluate this possibility, we tested the multiple logistic regression model presented above, with the inclusion of the number of services received by youths. Results of this regression showed that the number of services was positively correlated with recidivism. In other words, the more services received, the more likely the youth was to have an arrest during the follow-up period. This is most likely due to the fact that higher-risk youths are provided more services. In fact, the correlation between SDRRC resiliency and the number of services was -0.19 ($p < .0001$). We conducted supplemental analyses in which we divided the sample into low-, moderate-, and high-risk groups and performed the regression runs within each risk group. Results showed no significant relationship between the number of services and recidivism once youth resiliency was controlled for.

[back to top](#)

Discussion and Conclusions

Our analyses showed that the SDRRC has both internal and predictive validity for youth in the Los Angeles County Probation system. Total resiliency scores are correlated with total risk and total protective scores; subscales within risk and protective scores are significantly correlated with their overall scales. Subscales are often highly correlated with each other, however, suggesting a degree of redundancy in the instrument. The instrument and its subscales were significantly related to arrest for youths 12 months after their assessment. The scale was also significantly related to recidivism for major subgroups of interest: youths of different ethnicities, as well as both males and females. In analyses which took into account other factors related to recidivism, the SDRRC remained a significant predictor of subsequent arrest. However, the scale does seem to work differently for some youths. In particular, the scale is not as strong a predictor for Hispanic youths as for other youths.

Limitations of Current Research

Research studies are subject to limitations, and this one is no exception. Our follow-up was limited to 12 months following youth assessment with the SDRRC. Although this provides a window of time over which to observe behavior, longer follow-up time periods are preferable. Initially, a longer follow-up period had been planned, but the assessment phase took longer than expected.

As with many recidivism studies, our study relies on official records for measurement of youth behavior. We did not have access to youth’s self-reported criminal behavior, which can provide a more direct measure of criminal behavior (only a fraction of offenses result in arrest). Future research may want to examine the extent to which the SDRRC also corresponds with self-reported criminal behavior. To our advantage, however, the pilot test was conducted *before* the SDRRC was implemented. In this way, the validity testing was not contaminated by any system

policies or practices that were based on classifications by the SDRRC.

As indicated earlier, the SDRRC does not have any predetermined cut-points for resiliency. Without cut-points for classification, we could not conduct any meaningful analyses of false positives and false negatives—or the extent to which errors in prediction are made when using the SDRRC. Cut-points will be determined during the implementation phase of the instrument in Los Angeles. We recommend that sensitivity analyses be part of continued monitoring of the instrument once it has been integrated into Probation practices (as described below).

In addition, more thorough examination needs to be conducted on differences in the scales and subscales for different subgroups of youths. This should also be part of continued monitoring of the instrument.

Systemwide Implementation of LARRC

In summer of 2004, the Los Angeles County Probation Department started the process to institutionalize the SDRRC, now referred to as the LARRC. Training on LARRC began on August 4, 2004. In December 2004, staff began completing the LARRC assessment utilizing an automated system.

The Los Angeles County Probation Department has started a policy that requires all DPOs in the Juvenile Bureaus to assess and reassess minors assigned to their caseloads at defined intervals as part of a plan to enhance case management services. As investigators are trained in the administration of the LARRC, the assessment will be utilized at the investigation level (the point at which the pilot assessment was done) and will continue through the supervision stages in order to address protective/risk/resiliency factors, update case planning efforts, and link minors to appropriate services and interventions.

[back to top](#)

[References](#) | [Endnotes](#)

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

Table 1: Sampling Design for Validation Study

	Male		Female		
Ethnicity	Court	Non-court	Court	Non-court	Total
Black	178	89	89	44	400
Hispanic	178	89	89	44	400
White/other	178	89	89	44	400
Total	534	267	267	132	1200

Appendix

SAN DIEGO REGIONAL RISK & RESILIENCY ✓CHECKUP					DATE	PDJ #
					JAIN:	
YOUTH NAME (L/F/M)		NICKNAME	GENDER	AGE	DOB	HOME PHONE
RESIDENCE (STREET)			CITY	ZIP	ALT. PHONE (SPECIFY)	
SCHOOL		GRADE	ETHNICITY		INTERPRETER DESIRABLE <input type="checkbox"/> YOUTH <input type="checkbox"/> PARENT	
WHAT HAS ALREADY BEEN DONE FOR YOUTH/FAMILY?				CASE TYPE <input type="checkbox"/> COURT <input type="checkbox"/> NON-COURT		
MINOR: SPEAKS ENGLISH: <input type="checkbox"/> PROFICIENT <input type="checkbox"/> LIMITED <input type="checkbox"/> NONE		MINOR'S ASSESSMENT: PRIMARY LANGUAGE IN HOME:		MINOR'S PREFERRED LANGUAGE:		
PARENT/GUARDIAN: SPEAKS ENGLISH: <input type="checkbox"/> PROFICIENT <input type="checkbox"/> LIMITED <input type="checkbox"/> NONE		PARENT/GUARDIAN ASSESSMENT: PRIMARY LANGUAGE IN HOME:		PARENT/GUARDIAN PREFERRED LANGUAGE:		
PERSON COMPLETING THIS FORM:						
NAME		TITLE		OFFICE		PHONE
AGE AT FIRST ARREST				# PRIOR ARRESTS		

ADDITIONAL PROTECTIVE FACTORS						
1	(No) Commitment to School	0	1	2	3	
2	(No) Recognition for Involvement in Pro-social Activities	0	1	2	3	
3	(No) Relations with parents / other adults	0	1	2	3	
4	(No) Parental Monitoring	0	1	2	3	
5	(Negative) Parent Evaluation of Peers	0	1	2	3	
6	(No) Friends Engage in Conventional (Pro-social) Behavior	0	1	2	3	
7	(Not) Intolerant attitude towards deviance	0	1	2	3	
8	(No) Positive Social Orientation	0	1	2	3	
TOTAL ADDITIONAL PROTECTIVE SCORE			↓	↓	↓	=
			□	+	□	+
			□	+	□	=
			□	+	□	=

OTHER RISK FACTORS OR CONCERNS: (PLEASE CHECK ALL BOXES THAT MAY APPLY)	<input type="checkbox"/> ANIMAL CRUELTY <input type="checkbox"/> BLADDER CONTROL, DAYTIME LACK OF <input type="checkbox"/> BLADDER CONTROL, NIGHTTIME LACK OF <input type="checkbox"/> CHRONIC TARDINESS <input type="checkbox"/> EMOTIONAL DISTRESS <input type="checkbox"/> FIRE SETTING	<input type="checkbox"/> HEALTH PROBLEMS <input type="checkbox"/> HOMELESSNESS <input type="checkbox"/> INAPPROPRIATE SEXUAL BEHAVIOR <input type="checkbox"/> LOSS OR GRIEF <input type="checkbox"/> PARENTAL ABUSE/NEGLECT <input type="checkbox"/> PARENTAL REJECTION	<input type="checkbox"/> PEERS ARE OLDER/YOUNGER <input type="checkbox"/> PREDATORY OR HARASSING BEHAVIOR: <input type="checkbox"/> HATE CRIME <input type="checkbox"/> FOR PERSONAL GAIN <input type="checkbox"/> RACIALLY BASED <input type="checkbox"/> SEXUALLY BASED <input type="checkbox"/> SCHOOLYARD BULLYING	<input type="checkbox"/> SELF-MUTILATION <input type="checkbox"/> TOBACCO USE <input type="checkbox"/> VICTIM OF: <input type="checkbox"/> DOMESTIC VIOLENCE <input type="checkbox"/> PHYSICAL ABUSE <input type="checkbox"/> RACISM <input type="checkbox"/> SEXUAL, ABUSE
---	---	---	--	---

COMMENTS AND OBSERVATIONS:

SUMMARY SCORES			
TOTAL PROTECTIVE SCORE	□	TOTAL RESILIENCY SCORE	□
TOTAL RISK SCORE	□	TOTAL ADDITIONAL PROTECTIVE SCORE	□

INSTRUCTIONS: IF YOUR SELECTION IS NOT ABSOLUTELY AFFIRMATIVE, USE AN ARROW POINTING IN THE DIRECTION YOU WOULD LEAN TO IF GIVEN ANOTHER CHOICE. SEE EXAMPLE TO RIGHT.					EXAMPLE: YES SOMEWHAT NO UNK □ ■ → □ □							
DELINQUENCY - Protective					NET RISK SCORE	DELINQUENCY- Risk						
	YES	SOME WHAT	NO	UNK			YES	SOME WHAT	NO	UNK		
1	SUPPORT/REINFORCEMENT IN COMMUNITY	□	□	□		□	6	PRIOR ARRESTS	□	□	□	□
2	PRO-SOCIAL ADULT RELATIONS	□	□	□		□	7	SIGNIFICANT CRIME IN NEIGHBORHOOD	□	□	□	□
3	EXTENSIVE STRUCTURED ACTIVITIES	□	□	□		□	8	OFFENSES COMMITTED WHILE UNDER INFLUENCE	□	□	□	□
4	PARTICIPATES IN FAITH COMMUNITY	□	□	□		□	9	ASSAULTIVE OR FIGHTING BEHAVIOR	□	□	□	□
5	INVOLVED IN COMMUNITY ORGANIZATION	□	□	□	□	10	DELINQUENT ORIENTATION	□	□	□	□	
Delinquency Protective Subscale Score					+ → → →	Delinquency Risk Subscale Score						
EDUCATION - Protective						EDUCATION - Risk						
	YES	SOME WHAT	NO	UNK			YES	SOME WHAT	NO	UNK		
11	SCHOOL ENGAGEMENT/ BONDS	□	□	□		□	16	POOR ACADEMIC ACHIEVEMENT	□	□	□	□
12	ATTACHMENTS W/ACADEMIC ACHIEVER	□	□	□		□	17	PATTERN OF TRUANCY PAST YEAR	□	□	□	□
13	POSITIVE INTERACTIONS WITH TEACHERS	□	□	□		□	18	PATTERN OF SUSPENSION/EXPELLED	□	□	□	□
14	EDUCATIONAL ASPIRATIONS	□	□	□		□	19	DISRUPTIVE CLASSROOM/SCHOOL BX	□	□	□	□
15	CARING/SUPPORTIVE SCHOOL CLIMATE	□	□	□	□	20	PRESENTLY NOT IN EDUCATIONAL PROGRAM	□	□	□	□	
Education Protective Subscale Score					+ → → →	Education Risk Subscale Score						
FAMILY- Protective						FAMILY- Risk						
	YES	SOME WHAT	NO	UNK			YES	SOME WHAT	NO	UNK		
21	COMMUNICATES WITH FAMILY	□	□	□		□	26	POOR RELATIONS WITH PARENT(S)	□	□	□	□
22	CONSTRUCTIVE USE OF TIME AT HOME	□	□	□		□	27	PARENTAL SUPERVISION DEFICIENCIES	□	□	□	□
23	FAMILY ACTIVITIES	□	□	□		□	28	CHAOTIC FAMILY	□	□	□	□
24	FAMILY SUPPORT	□	□	□		□	29	PARENTAL CRIMINALITY/ SUBSTANCE ABUSE	□	□	□	□
25	UNCONDITIONAL REGARD FROM A PARENT	□	□	□	□	30	RUNAWAY	□	□	□	□	
Family Protective Subscale Score					+ → → →	Family Risk Subscale Score						
PEER - Protective						PEER - Risk						
	YES	SOME WHAT	NO	UNK			YES	SOME WHAT	NO	UNK		
31	POSITIVE PEER RELATIONS	□	□	□		□	36	SOCIALLY ISOLATED	□	□	□	□
32	HAS AT LEAST ONE PERSON TO CONFIDE IN	□	□	□		□	37	VERY FEW PROSOCIAL ACQUAINTANCES	□	□	□	□
33	VALUES DIGNITY/RIGHTS OF OTHERS	□	□	□		□	38	HAS GANG AFFILIATION/ASSOCIATION	□	□	□	□
34	ABILITY TO MAKE FRIENDS	□	□	□		□	39	HAS DELINQUENT FRIENDS	□	□	□	□
35	ABILITY TO COMMUNICATE DISAGREEMENTS	□	□	□	□	40	NO MEANINGFUL RELATIONSHIP W/ANY ADULT	□	□	□	□	
Peer Protective Subscale Score					+ → → →	Peer Risk Subscale Score						
SUBSTANCE USE- Protective						SUBSTANCE USE - Risk						
	YES	SOME WHAT	NO	UNK			YES	SOME WHAT	NO	UNK		
41	PARENTS MODEL HEALTHY MODERATION	□	□	□		□	46	PATTERN OF ALCOHOL USE	□	□	□	□
42	EFFECTIVELY MANAGES PEER PRESSURE	□	□	□		□	47	USED MOOD ALTERING SUBST. (OTHER THAN ALCOHOL)	□	□	□	□
43	YOUTH IS FREE OF DISTRESSING HABITS	□	□	□		□	48	USES SUBSTANCES FREQUENTLY	□	□	□	□
44	YOUTH MANAGES STRESS WELL	□	□	□		□	49	SUBSTANCE USE INTERFERES WITH DAILY FUNCTIONING	□	□	□	□
45	POSITIVE SELF-CONCEPT	□	□	□	□	50	EARLY ONSET SUBSTANCE USE (<13)	□	□	□	□	
Substance Use Protective Subscale Score					+ → → →	Substance Use Risk Subscale Score						
INDIVIDUAL - Protective						INDIVIDUAL - Risk						
	YES	SOME WHAT	NO	UNK			YES	SOME WHAT	NO	UNK		
51	VALUES HONESTY/INTEGRITY	□	□	□		□	56	NO PROSOCIAL INTERESTS	□	□	□	□
52	SELF CONTROL	□	□	□		□	57	SUPPORTIVE OF DELINQUENCY	□	□	□	□
53	SELF EFFICACY IN PROSOCIAL ROLES	□	□	□		□	58	ANGER MANAGEMENT ISSUES	□	□	□	□
54	PROBLEM-SOLVING SKILLS	□	□	□		□	59	SENSATION SEEKING	□	□	□	□
55	PLANS, ORGANIZES, & COMPLETES TASKS	□	□	□	□	60	MANIPULATIVE/DECEITFUL	□	□	□	□	
Individual Protective Subscale Score					+ → → →	Individual Risk Subscale Score						
TOTAL PROTECTIVE SCORE					TOTAL RESILIENCY SCORE	TOTAL RISK SCORE						

Table 2: Demographic Characteristics (Unweighted)

		N (%)
Gender		
	Male	768 (74.1%)
	Female	268 (25.9%)
Age		
	9-12	65 (6.3%)
	13-14	240 (23.2%)
	15-16	404 (39.0%)
	17-18	322 (31.1%)
	19+	5 (0.5%)
Ethnicity		
	White	194 (18.7%)
	Black	299 (28.9%)
	Hispanic	436 (42.1%)
	Other	97 (9.4%)
	Unknown	10 (1.0%)
Case Type		
	Court	782 (75.5%)
	Non-court	254 (24.5%)
	Investigation	294 (28.4%)
	Supervision	742 (71.6%)

Table 3: Mean Resiliency Scores by Demographic Characteristics (Weighted)

		Resiliency Score	Number in Sample
Gender			
	Male	18.9	768
	Female	22.1	268
Age			
	9-12	23.1	65
	13-14	20.5	240
	15-16	17.7	404
	17-18	20.9	322
	19+	-0.3	5
Ethnicity			
	White	25.7*	194
	Black	21.5*	299
	Hispanic	16.0*	436
	Other	32.6*	97
	Unknown	12.1	10

* $p < .05$ using t-tests

Table 4: Correlations Between Total Resiliency Score and Subscale Items (Weighted)

Score	Correlation
Total protective score	0.93
Total risk score	0.88
Net risk for delinquency	0.85
Net risk for education	0.81
Net risk for family	0.81
Net risk for peer	0.87
Net risk for substance use	0.81
Net risk for individual	0.88
Delinquency risk factors	0.64
Education risk factors	0.68
Family risk factors	0.60
Peer risk factors	0.70
Substance use risk factors	0.54
Individual risk factors	0.73
Delinquency protective factors	0.75
education protective factors	0.78
Family protective factors	0.81
Peer protective factors	0.77
Substance use protective factors	0.77
Individual protective factors	0.82

Note: All correlations significantly different from zero ($p < .05$).

Table 5: Arrested Within 12 Months of Assessment, by Resiliency Score (Weighted)

	Total	Delinquency	Education	Family	Peer	Substance use	Individual
Total	1.00	0.85	0.82	0.82	0.88	0.82	0.88
Delinquency		1.00	0.64	0.64	0.69	0.65	0.71
Education			1.00	0.62	0.66	0.54	0.60
Family				1.00	0.68	0.60	0.78
Peer					1.00	0.65	0.78
Substance use						1.00	0.70
Individual							1.00

Note: All correlations significantly different from zero ($p < .05$).

Table 6: Arrested Within 12 Months of Assessment, by Resiliency Score (Weighted)

Resiliency Score	No	Yes	% of sample
Low (12 or less)	64.5%	35.5%	35.8%
medium (13-33)	84.5%	15.5%	33.6%
High (34+)	91.8%	8.2%	30.6%
Total	79.6%	20.4%	100.0%

Chi-square = 88.3 (p < .0001) with 2 degrees of freedom

Figure 1: Percent Arrested During Follow-Up, by Age and Resiliency Score (Weighted)

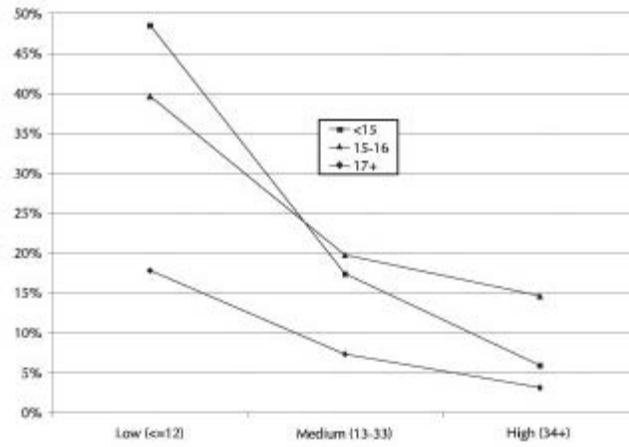


Figure 2: Percent Arrested During Follow-Up, by Gender and Resiliency Score (Weighted)

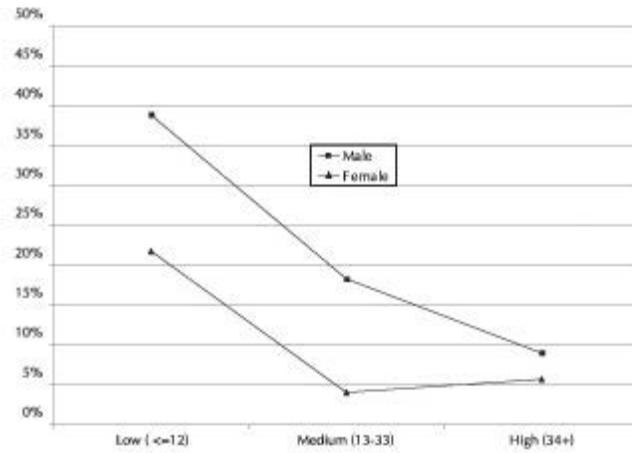


Figure 3: Percent Arrested During Follow-Up, by Ethnicity and Resiliency Score (Weighted)

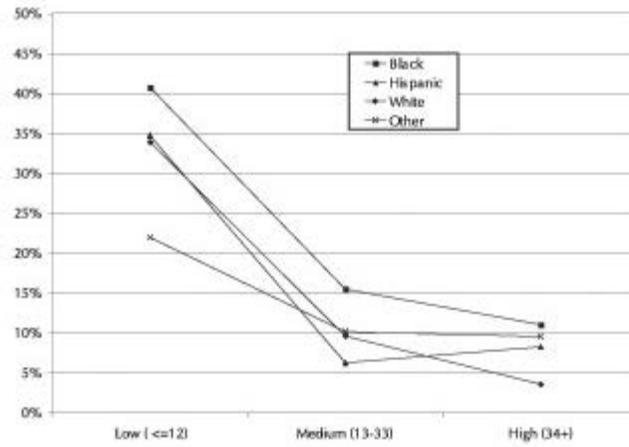


Table 7: Mean Assessment Scores and Correlations with Arrest During Follow-Up (Weighted)

Score	Mean	Correlation
Total resiliency	19.55	-0.27
Total protective	33.72	-0.25
Total risk	-14.17	-0.24
Net risk for delinquency	1.71	-0.24
Net risk for education	2.02	-0.24
Net risk for family	5.20	-0.19
Net risk for peer	4.31	-0.24
Net risk for substance use	3.89	-0.19
Net risk for individual	2.42	-0.23
Delinquency risk factors	-2.84	-0.21
Education risk factors	-3.25	-0.21
Family risk factors	-1.75	-0.13
Peer risk factors	-2.10	-0.19
Substance use risk factors	-1.65	-0.12
Individual risk factors	-2.58	-0.19
Delinquency protective factors	4.55	-0.19
Education protective factors	5.27	-0.23
Family protective factors	6.94	-0.20
Peer protective factors	6.42	-0.21
Substance use protective factors	5.54	-0.19
Individual protective factors	5.00	-0.21

Note: All correlations in this table are significantly greater than zero ($p < .05$).

Table 8: Logistic Regression Results for Arrest During Follow-Up (Weighted)

Variable	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Sq
Intercept	-20.4302	6.7603	9.1329	0.0025
Age	2.6623	0.9056	8.6421	0.0033
Age squared	-0.0960	0.0301	10.1707	0.0014
Male	0.9814	0.2408	16.6081	<.0001
Black	0.1976	0.2046	0.9325	0.3342
White	-0.2881	0.2680	1.1556	0.2824
Other race	-0.1627	0.3532	0.2121	0.6451
Supervision	1.5024	0.4206	12.7614	0.0004
Court case	-0.7125	0.3766	3.5786	0.0585
Resiliency	-0.0285	0.00430	43.8683	<.0001

Table 9: Logistic Regression Results for Arrest During Follow-Up, With Interaction Terms (Weighted)

Variable	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Sq
Intercept	-20.0157	6.6641	9.0212	0.0027
Age	2.5887	0.8937	8.3909	0.0038
Age squared	-0.0935	0.0297	9.9074	0.0016
Male	1.0138	0.2447	17.1713	<.0001
Black	0.4180	0.2456	2.8975	0.0887
White	0.0476	0.3186	0.0223	0.8812
Other race	-0.2739	0.4633	0.3496	0.5544
Supervision	1.5069	0.4182	12.9845	0.0003
Court case	-0.6778	0.3752	3.2631	0.0709
Resiliency	-0.0217	0.00551	15.4913	<.0001
Resiliency* Black	-0.0171	0.0103	2.7451	0.0976
Resiliency* White	-0.0313	0.0141	4.9051	0.0268
Resiliency* Other	0.000710	0.0140	0.0026	0.9596

[Home](#)

How Much Risk Can We Take? The Misuse of Risk Assessment in Corrections

[Endnotes](#)

James Austin, Ph.D.
The JFA Institute

[The Basics of Risk Assessment](#)
[A Closer Look at the LSI-R](#)
[Linking Risk to Punishment and Treatment?](#)

AFTER DECADES OF intellectual neglect, the field of corrections has decided to embrace the world of science and adhere to the dictums of “evidence-based” corrections. The term “evidence based” originates from the field of medicine as far back as the 19th century in Europe and means many things to many people. ^[1] In medicine it is very important that medical procedures and the use of healing drugs and medicine actually demonstrate their effectiveness through rigorous experimental studies before they are brought to market. In the social sciences, evidence-based research suggests that governmental policies must be shaped by scientific evidence that shows the policy has some cause and effect value. For many good reasons, the field of corrections has never had to pass such a high standard. But after American corrections has set world records in the numbers of persons incarcerated and placed on probation and parole, some criminal justice professionals believe the field needs to get serious about its \$60 billion a year industry and produce a better product.

Plagued by recidivism rates that have remained stubbornly stagnant for 30 years (or more) and by a general feeling among most politicians that about the only thing that corrections can do is inflict widespread punishment, criminal justice practitioners have seen the more benign goals of treatment and rehabilitation take a back seat to the more politically appealing ideologies of deterrence, incapacitation, and retribution. It’s a given that no politician can successfully run on a platform demanding more and better treatment for the two million plus prisoners held in our nation’s jails and prisons.

But the times are a changing. Led by a small number of Canadian and American criminologists, there is now a considerable effort to get rehabilitation and treatment back on the map. Their argument is advertised not as ideological but as empirical. The major premise is that treatment does work if it is done right. Therefore, the primary reason treatment is ineffective is because it is more often done wrong.

One major reason that treatment is not done right is that offenders are not properly assessed for risk by most correctional agencies. Without the proper diagnosis, it is not possible to assign prisoners to the proper treatment. Indeed, prior research has shown that assigning low-risk people to treatment they really don’t need actually increases recidivism. A recent evaluation of Ohio’s community corrections act clearly shows that many correctional programs are not targeting the proper offender, which in turn diminishes the capacity to reduce recidivism rates. ^[2]

The widespread absence of risk assessment in corrections has historically hampered correctly targeted treatment. It was not until the 1980s that prison systems, due in part to a number of federal lawsuits, finally started using custody classification systems to assign prisoners to the correct prisons. The results have been impressive in most states, with increasing numbers of prisoners now being assigned to minimum security settings. The taxpayers have benefited somewhat because the lower the security, the lower the incarceration costs. Unfortunately, the huge increases in the correctional populations have largely negated whatever savings taxpayers would have realized.

Parole boards, which still govern the date and conditions of release for prisoners in most states, have only recently (and only in a few states) embraced the idea that their decisions would be influenced by some calculation of the prisoner's risk to recidivate. Probation and parole agencies have also begun to implement risk instruments to guide their decisions as to what levels of supervision are most appropriate for their burgeoning caseloads.

But despite these advances, no jurisdiction can point to significant reductions in recidivism rates—and that includes Canada, from which most of the new emphasis on rehabilitation has emanated. ³ Many probation and parole officers seem less interested in risk assessment and case management and more concerned with racking up as many violations of their caseload as they can. I don't recall any prison, parole, or probation department being chastised for having too high a recidivism rate, even though there is considerable evidence that they could have a positive effect on these rates.

The remainder of this article focuses on the state of risk assessment. I concede that in order for rehabilitation to have a meaningful impact on recidivism rates, the proper identification of persons by their risk level is essential. But I now worry that the field is placing too much emphasis on risk assessment with little effort to provide those basic treatment services that are needed.

[back to top](#)

The Basics of Risk Assessment

Before an agency decides to adopt a risk assessment system, a number of tests need to be completed to ensure it will work. There seems to be a trend in corrections to uncritically accept the latest "innovation" and adopt it without understanding its strengths and limitations. In risk assessment, unless these steps are completed, application of the risk assessment process may prove more harmful than helpful as offenders will be improperly classified.

1. Risk Assessment Instruments Must Be Tested on Your Correctional Population and Separately Normed for Males and Females.

There is a tendency for correctional agencies to simply borrow or buy an instrument that has been developed on another population that may not reflect the attributes of their own offender populations. In research terms this issue has to do with the "external validity" of the instrument and the ability to generalize the findings of a single study of the instrument to other jurisdictions. Generally, if a risk assessment instrument has not been tested on multiple populations under varying conditions, it will not work well on populations it has not been tested on.

Male and female risk assessment is another issue for proposed risk instruments. Men and women are different, behave differently, and respond differently to various forms of treatment and supervision. Yet when it comes to risk assessment we often assume they are the same. Recidivism and career criminal studies consistently show that females are less involved in criminal behavior, are less likely to commit violent crimes and are less likely to recidivate after being placed on probation or parole. Further, since the "criminal population" is largely male, any instrument that is tested on a total correctional population will naturally misclassify females.

2. An Inter-Rater Reliability Test Must Be Conducted

Both an inter-reliability test and a validity test must be completed by independent researchers who have no economic gain in proving the effectiveness of the instrument. Inter-rater reliability has to do with the accuracy and consistency of the instrument being completed by those who will be responsible on a day-to-day basis for completing the form and interpreting the results. Often this work is done by probation and parole officers or parole board hearing examiners. It is a skilled task that not all correctional staff are well suited for.

The inter-rater reliability test would consist of taking a representative sample of offenders (a minimum of 100 cases) who will then be independently scored using the proposed instrument by two staff who have trained in the proposed instrument. Any item on the instrument that does not reach the 80 percent agreement level should be deleted. If the instrument does not demonstrate an agreement level of 90 percent, it should not be implemented.

3. A Validity Test Must Be Conducted

The validity test is designed to see how well the risk factors actually predict recidivism. This test is done by drawing a sample of offenders who were sentenced to probation or released from prison and tracking them for a period of 2 to 3 years. Since most jurisdictions are anxious to have the risk assessment instrument implemented as quickly as possible, the validation sample often consists of persons sentenced or released 2 to 3 years prior to the study being conducted. The research must then be able to perform a variety of bi-variate and multi-variate statistical tests to determine which items should be used, the weights assigned to each item and the proper risk level scale.

4. The Instruments Must Allow For Dynamic and Static Factors That Have Been Well Accepted and Tested In A Number of Jurisdictions

As noted above, the risk instrument should consist of static and dynamic risk items. [Table 5](#) summarizes commonly used risk factors that have been repeatedly validated by a number of validation studies. These are separated into the static and dynamic categories. Of the two, the dynamic factors are generally the more powerful predictors, as they reflect the person's current social and economic environment. If an instrument does not employ dynamic factors, it is likely to not perform accurately.

5. The Instruments Must Be Compatible with the Skill Level of Your Staff

There are a wide variety of risk assessment instruments available to jurisdictions to use. However, they require very different skill levels. The more traditional risk assessment forms generally consist of not more than 10-12 items and are based on factual items that can be gleaned from court and case files and require minimal interpretation by staff trained in their use. Age at first arrest, current age, and number of prior probation violations within the past five years come under this category. For these instruments staff need little academic training to conduct an accurate assessment.

The more complicated risk assessment items require a well-structured interview and a review of all relevant case file data. These instruments often have 40-60 items with several sub-scales reflecting varying domain risk levels. With such instruments it is more difficult to achieve the minimal levels of reliability and validity, unless the staff is highly skilled in the application of psychometric assessment forms. Without such skilled staff, the use of these instruments is not recommended.

6. The Risk Assessment Must Have "Face Validity" And Transparency with Staff, Prisoners, Probationers, Parolees and Policy Makers.

Finally, the instrument and the entire risk assessment process needs to be credible with all of the parties that are being directly impacted by it. Staff assigned to the risk assessment process must believe that the instrument actually works and will help inform the decision process for sentencing, release, and supervision decisions. The decision makers (judges, parole boards, and

correctional administrators) must also have confidence in the risk assessment process and demonstrate through their decisions that they are using it. In particular, statistics should show that offenders assessed as low risk should have lower rates of being sentenced to prison, have shorter sentences, have high rates of being paroled and receive lower levels of supervision. High-risk offenders should show just the opposite trends.

The people who are being assessed for risk must also believe that the process is credible and will be used by decision makers. The process should also be transparent and not some mysterious process where the offender is unaware of what factors are being used and how each is scored. This is especially helpful for risk instruments that employ dynamic risk factors—items that can change based on the offender’s social and economic situation (employment, residency, and family relations). By understanding these dynamic risk factors, the offender can take actions or seek support that will actually reduce the risk to public safety.

[back to top](#)

A Closer Look at the LSI-R

As the interest in risk assessment has grown, so too has the private industry engaged in developing and distributing these systems. Currently there are two major privately held risk assessment systems available to corrections. The most widely advertised system is the Level of Service Inventory—Revised (or LSI-R), which was first developed in Canada and has now been adopted by a number of U.S. correctional agencies. ⁴ LSI-R is owned and distributed by Multi Health Systems, Incorporated, which distributes a wide array of psychologically based assessment tools. ⁵ The other is the Correctional Offender Management Profiling for Alternative Sanctions (or Compas), owned and distributed by the Northpointe Institute for Public Management, Inc., which also offers a privately held prison and jail classification system. ⁶

Few independent validation studies of these two systems appear in the literature. By “independent” I mean studies done by researchers who have no financial interest in the two companies. Because the LSI-R has been around longer and is more widespread than the Compas, there have been a few recent studies in Washington, Pennsylvania, and now Vermont. As will be shown below, these studies show that many of the individual factors used in the LSI-R scale are not predictive of re-offending behavior. ⁷

Why is this so? The principal problem with the LSI-R is that it is difficult to achieve a sufficient level of inter-rater reliability on many of its items. The LSI-R consists of 54 items that are sorted into the following ten substantive areas believed to be related to future criminal behavior:

1. Criminal History (10 items)
2. Education and Employment (10 items)
3. Financial (2 items)
4. Family and Marital (4 items)
5. Accommodations (3 items)
6. Leisure and Recreation (2 items)
7. Companions (5 items)
8. Alcohol and Drugs (9 items)
9. Emotional and Personal (5 items)
10. Attitude and Orientation (4 items)

The LSI-R scorer is expected to make either a dichotomous “yes” or “no” to 37 items and a likert scale rating of satisfactory, relatively satisfactory, relatively unsatisfactory or very unsatisfactory for the other 17 items. For example, one question in the family/marital domain requiring a level of satisfactory response is “dissatisfaction with marital or equivalent situation.” The scorer is instructed to base this assessment on a review of the case file data and an interview with the subject. On the accommodation domain, one question requiring a yes/no response is “three or more address changes last year.” Such questions and the associated response raise important questions about whether correctional staff (most of whom have little if any

training in psychometric testing) can correctly use this assessment.

Researchers in Washington State conducted one of the first independent validation studies of the LSI-R as it was being applied to released state prisoners. ⁸ The authors found that the LSI-R criminal history factors were strong predictors of recidivism and produced most of the predictive power for the instrument. Put differently, many of the numerous other LSI-R items do little to enhance the LSI-R predictive attributes. These findings led the researchers to recommend that some of the LSI-R items be combined with other non-LSI-R factors, like current age and gender, to provide for a better risk instrument.

A recent study of the LSI-R as used by the Pennsylvania Parole Board and the Department of Corrections is instructive on difficulties associated with the LSI-R scoring process. ⁹ In particular, it provides the results of an inter-rater reliability study—a study that should be done for any risk assessment system. The Pennsylvania Parole Board was using the LSI-R scores to determine the suitability for release from prison. However, there had been no attempt to validate the LSI-R on Pennsylvania prisoners, which no doubt are somewhat different from the Canadian prisoners on whom the LSI-R had been developed. Further, the concept of using LSI-R for parole release considerations suggests a serious misapplication of the LSI-R, since many of the items have to do with the prisoner’s life prior to incarceration. For example, how does one assess whether a prisoner who has been incarcerated for several years has “some criminal acquaintances” or few anti-criminal friends”? Given that many months or years may have passed since the offender was living in the community, the problems of accurate recall and the relevance of the questions for prisoners are rather obvious.

But even with these issues, one must also determine if the assessors are able to produce reliable scoring results. To this end, several reliability tests were conducted. The basic test is relatively straightforward and easy to do. A sample of 120 cases was selected for the test. Within two weeks, two staff were required to independently score the sampled cases and determine the appropriate score for each case. The results are shown in [Table 1](#). The table contains only the 16 items that reached the 80 percent level of inter-rater agreement. The other 38 items had scores in the 60-70 percent range. If we use the more generous criteria of risk level, the level of disparity is reduced but remains at an unacceptable range, with a 29 percent disagreement on the risk level. It is also noteworthy that the items that have an acceptable reliability score are the more factual ones that are found in the more traditional risk assessment instruments.

With such a level of “noise” in the scoring process, it is not surprising that only a few of the LSI-R items were found to be associated with recidivism. A recidivism study of 1,006 prisoners who were scored on the LSI-R and had been released for at least one year was conducted. The first task was to perform an item by item test of 54 LSI-R scoring items to see which ones were associated with recidivism. This analysis showed that only the following items had a statistical association with recidivism:

1. Any prior convictions?
2. Two or more prior convictions?
3. Three or more prior convictions?
4. Arrested under age 16?
5. Escape History?
6. Probation/parole suspension during prior community supervision?
7. Three or more address changes the past year?
8. Current drug problem?
9. Drug problem related to law violations?
10. Drug problem related to school or work problems?
11. Mental health problems in the past?

[Table 2](#)

A regression analysis was done to see which of these 11 items had an independent effect on recidivism. This resulted in the following eight items being used: any prior convictions, two or

more prior convictions, arrested under age 16, prior probation/parole suspension, three or more address changes within the last year, current drug problem, problem affecting school/work, and mental health treatment in the past.

As shown in [Table 3](#), only a small number of the 54 LSI-R scoring items are useful and most of them are not contributing to the risk assessment process. We also found that compared to the risk groups created by the full LSI-R, the condensed instrument creates risk categories with greater distinctiveness in terms of recidivism. Not only do these items have better predictive ability, but also they reduce the “high risk” category. According to this instrument, only 188 prisoners would be classified as “high risk,” compared to 522 using the full LSI-R instrument. More importantly, the high-risk group created by the condensed instrument has a 69 percent recidivism rate, compared to the 58 percent recidivism rate of the LSI-R high risk group, indicating that the condensed instrument does a better job of selecting those prisoners representing the most significant danger to public safety.

In [Table 4](#), the analysis is taken a step further. Along with the eight LSI-R items in the condensed instrument, we also include these descriptive variables: age at release, marital status, committing offense, and release type. This instrument, combining a small number of reliable LSI-R items with a few demographic items, produced the best risk assessment results. In this analysis, we are able to develop greater specificity within the “low risk” category and to identify groups of prisoners with more distinct rates of re-offending.

In Vermont a similar set of findings were noted. The study was similar to the Pennsylvania one where the Parole Board was desirous of adopting a risk instrument to guide their decision-making process. Although a formal reliability was not completed, a validation study was made on the LSI-R and other factors believed to associate with risk to public safety. Two measures of recidivism rates were tested (return to prison and a new conviction), and became the basis for determining which items should be included in the final risk assessment instrument. The first step was to conduct a bi-variate statistical analysis to determine which items had a simple association with the three measures of recidivism.

The original validation study was based on 2,533 sentenced prisoners who were released in 2002. Of this number only 644 had completed LSI-R scoring results. As was found in the Washington state and Pennsylvania studies, only a relatively small number (13) of the LSI-R 54 factors are consistent and strong predictors of recidivism (items 4, 8, 9, 11, 13, 14, 16, 17, 31, 34, 39, 40 and 50). And another set of variables that are not part of the LSI-R was found to be associated with recidivism rates. These included current age, marital status, education level, measures of institutional conduct, and completion of certain programs while incarcerated (see [Table 5](#)).

[back to top](#)

Linking Risk to Punishment and Treatment?

The above studies show that risk assessment is doable but that it need not be complicated or expensive. Before adopting a particular system, an agency needs to rigorously assess what model it can afford and administer in a professional and accurate manner. If the wrong decisions are made in terms of what model to buy you may end up with little if any enhancements to your ability to assess risk.

I want to close on another matter that seems to be receiving little attention; namely, the requirement to administer or provide the proper “intervention” that is consistent with risk. The major assumption in evidence-based policy is that prisoners, probationers and parolees are to be “serviced” and punished relative to their risk. But reaching this standard can fail for two reasons. First, the assessed risk level becomes moot if there are no high quality programs or interventions to assign the “client” to once the assessment has been completed. For example, in Vermont only 14 percent of the released prisoner sample had completed an educational, substance abuse, or sex treatment program while incarcerated, even though 31 percent of the sample were assessed as high risk.

On the other end of the spectrum, we need to recognize that a very large proportion of the prison, probation, and parole populations is low risk; these offenders are being punished and even treated beyond their threat to public safety. It's like a hospital that decides to provide intensive care for patients who have a cold—the treatment is not only unnecessary but expensive.

It would be helpful for those in the risk assessment business to start advocating a more reasonable level of intervention that matches the risk they have so carefully calibrated.

[back to top](#)

[Endnotes](#)

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

Table 1: LSI-R Reliability Scores at the 80% Level

Variable	% Agreement
1. Any prior convictions?	96%
2. Two or more prior convictions?	93%
3. Three or more convictions?	93%
4. Three or more present offenses?	81%
6. Ever incarcerated upon conviction?	95%
7. Escape history from a correctional facility?	81%
8. Ever punished for institutional misconduct?	87%
9. Charge/probation/parole suspended during prior community supervision?	91%
10. Official record of assault/violence?	86%
11. Currently employed?	86%
12. Less than regular grade 10?	85%
13. Less than regular grade 12?	88%
14. Three or more address changes last year?	82%
15. Drug problem, ever?	88%
16. Moderate interference?	84%
17. Severe interference, active psychosis?	93%
18. Mental health treatment, past?	87%
19. Mental health treatment, present?	89%

Table 2: Cross-tabulation of the First and Second LSI-R Interviews of the Reliability Sample

1st Interview	2nd Interview			
	Low	Medium	High	Total*
Low (0 through 15)	3	2	0	5
medium (16 through 22)	3	14	11	28
High (23 and above)	1	17	67	85
Total	7	33	78	118

*Note: Two cases were not scored a second time and were excluded from this analysis.

**Table 3: Score using Select LSI-R Items by Failure Variable
Pennsylvania Board of Probation and Parole**

Point Distribution		Total		Recidivated
		N	%	%
TOTAL		948	100.0%	53.4%
Low	0	17	1.8%	17.6%
	1	17	1.8%	29.4%
	2	58	6.1%	22.4%
	3	54	5.7%	33.3%
	Total Low Risk	146	15.4%	33.6%
Moderate	4	119	12.6%	52.9%
	5	115	12.1%	47.8%
	6	133	14.0%	51.9%
	7	160	16.9%	56.3%
	8	87	9.2%	58.6%
	Total Moderate Risk	614	64.8%	53.4%
High	9	115	12.1%	70.4%
	10	23	2.4%	73.9%
	11	43	4.5%	60.5%
	12	1	0.1%	0.0%
	13	6	0.6%	83.3%
	Total High Risk	188	19.8%	68.6%

Table 4: Score using Select LSI-R and Demographic Items by Failure

Pennsylvania Board of Probation and Parole				
Point Distribution		Total		Recidivated
		N	%	%
TOTAL		848	100.0%	53.2%
Lowest	1	1	0.1%	0.0%
	2	7	0.8%	0.0%
	3	23	2.7%	21.7%
	Total Lowewt Risk		31	3.7%
Low	4	20	2.4%	45.0%
	5	43	5.1%	37.2%
	6	64	7.5%	37.5%
	Total Low Risk		127	15.0%
Moderate	7	89	10.5%	51.7%
	8	91	10.7%	46.2%
	9	115	13.6%	52.2%
	10	104	12.3%	57.7%
	11	94	11.1%	56.4%
	Total Moderate Risk		493	58.1%
High	12	92	10.8%	65.0%
	13	41	4.8%	78.0%
	14	29	3.4%	89.7%
	15	19	2.2%	52.6%
	16	12	1.4%	66.7%
	17	4	0.5%	75.0%
	Total High Risk		197	23.2%

Table 5: Vermont Parole Board Final Risk Assessment Simulation Score by Item and Overall Risk Level

Static Item		N-644	%	% Reconvicted
1. Age at First Arrest*				
	16 years or older	458	71%	45%
	Under 16 years	186	29%	60%
2. Prior Charges/Suspensions under Supervision*				
	No	169	26%	36%
	Yes	475	74%	54%
3. Crime Seriousness				
	Level 1,2,4,5 and 10	334	52%	37%
	Level 3,6,7,8,9 and 11	310	48%	61%
4. Drug/Alcohol Abuse*				
	None	388	60%	42%
	Behavior problems/Positive test	256	40%	60%
5. Prior Convictions*				
	None	387	60%	45%
	One	109	17%	49%
	Two or more	148	23%	60%
6. Criminal Acquaintances at Admission*				
	No	43	7%	26%
	Yes	601	93%	51%
7. Employed at least 12 months prior to Admission*				
	Yes	221	34%	41%
	No	423	66%	53%
Dynamic Item		N-644	%	% Reconvicted
8. Current Age				
	50 and above	46	7%	20%
	40-49	133	21%	33%
	24-39	325	50%	55%
	Under 23	140	22%	61%

9. Most Severe Disciplinary report				
	None in the past 12 months	450	70%	44%
	major A or major b	194	30%	59%
10. Completed education/Substance Abuse program				
	Yes	87	14%	45%
	No	557	86%	50%
11. Current Custody Level				
	Minimum	0	0%	N/A
	Medium	0	0%	N/A
	Else	0	0%	N/A
12. Current marital Status				
	Married, Divorced	204	32%	43%
	Single	440	68%	52%
Scored Risk Level				
	Low	151	23%	26%
	Moderate	293	45%	49%
	High	200	31%	67%

*Denotes LSI-r factors

[Home](#)

Assessing the Role of Clinical and Actuarial Risk Assessment in an Evidence-Based Community Corrections System: Issues to Consider

[References](#)

James M. Byrne

*Department of Criminal Justice and Criminology, University of Massachusetts, Lowell
April Pattavina*

Department of Criminal Justice and Criminology, University of Massachusetts, Lowell

[Issue 1: Evidence-based Practice and the \(Missing\) Link Between Risk Assessment and Risk Reduction](#)

[Issue 2: The Implications of Actuarial and Clinical Assessment for Line Staff and Management](#)

[Issue 3: The Need to Combine Individual and Community Risk Assessment](#)

[Concluding Comments](#)

THE RISK ASSESSMENT process is undergoing major change in federal, state and local community corrections agencies across the country. New assessment instruments are being introduced, case management systems are being redesigned, and the roles and responsibilities of line staff and management in community corrections agencies are being redefined, in large part due to the application of new, “soft” computer technology in community corrections agencies (Pattavina and Taxman, 2006). As Gottfredson and Tonry (1987) predicted in the late 1980s, “both the literature and practical application of science-based prediction and classification will continue to expand as institutions evolve to become more rational, more efficient, and more just” (vii). While rationality, efficiency, and justice are laudable goals for any criminal justice organization, we suspect that ultimately, it is the effectiveness of the community corrections system—both in terms of short-term offender control and long-term offender change—that really matters to the public, and by extension, to policymakers and practitioners. In the following article, we examine three key issues related to assessing the effectiveness of risk assessment procedures that need to be addressed: 1) evidence-based practice and link between risk assessment and risk reduction, 2) the implications of both actuarial and clinical assessment for line staff and management, and 3) the need to combine individual risk assessment and community risk assessment in the next generation of risk-driven community corrections strategies. We conclude by offering three simple recommendations designed to improve the effectiveness of the risk assessment process in federal, state, and local community corrections agencies.

[back to top](#)

Issue 1: Evidence-based Practice and the (Missing) Link Between Risk Assessment and Risk Reduction

When the term “best practices” is used, it typically refers to the results of an evidence-based

review of the research on a topic of interest (e.g. scared straight programs, prison and community-based treatment programs, etc.). Essentially, there are three different types of evidence-based reviews: 1) the “gold standard” evidence-based review focuses only on randomized, controlled experiments; 2) the “bronze standard” evidence-based review includes both experimental and well-designed quasi-experimental research, while using nonexperimental research studies to confirm findings from higher quality research; and 3) the unscientific (or nonsense) review, which does not identify specific study review criteria, relying instead on a selected subset of all studies available for review on the topic of interest. Not surprisingly, the last category of unscientific reviews is usually written by advocates of a particular program or strategy. In their most extreme form, the authors of the review simply allude to an evidence-based review or to “best practices,” with no supporting documentation. Unfortunately, much of what is currently available in the corrections field—both institutional and community corrections—falls into this last category.

Applying the “gold standard” for evidence-based reviews to the “risk assessment” process in community corrections mandates that at least two randomized field experiments must have been conducted in this area before we can offer an assessment of “what works” (see, for example, the reviews conducted for the Campbell/Cochrane collaboration at www.campbellcollaboration.org). Unfortunately, no experimental research has been conducted on this topic in community corrections, leading us to conclude that we don’t know whether there is a link between risk *assessment* (i.e., classification of an offender into high-risk, medium-risk and low-risk classification categories) and risk *reduction* (i.e., a lower rate of recidivism for offenders than anticipated, given their risk level) due to the types of supervision and services we make available to offenders at each level of risk.

Much of what we currently do in community corrections is based on assumptions about the risk-reduction effects of placing offenders into different supervision levels that have not been tested empirically, using randomized field experiments. What would happen, for example, if we placed high-risk offenders under “medium” or “low” supervision? Alternatively, what would be the impact of placing a low- or medium-risk offender under “maximum” supervision? Until we have the results of quality, experimental research to review, we will continue to make assumptions about “what works and “best practices” in terms of both risk assessment and risk reduction that are simply not supported by a careful, “gold-standard” evidence- based review.

[back to top](#)

Issue 2: The Implications of Actuarial and Clinical Assessment for Line Staff and Management

One argument that can be made concerning the use of clinical vs. actuarial risk assessment is that the line staff currently hired in community corrections do not have the background and qualifications necessary to conduct “clinical” assessments of offender risk, particularly for special category (e.g. mental health, substance abuse, sex offender) and multiple problem offenders. Assuming for the sake of argument that you want to introduce clinical assessments into your federal, state, and local community corrections agency, you have two choices: 1) recruit/hire line staff with the necessary qualifications to conduct clinical risk assessments (perhaps with minimal additional training); or 2) privatize the assessment process, using the network of current mental health treatment providers as your “target” potential provider. Indeed, it could be argued that by moving away from clinical and toward actuarially-based risk assessments, we are attempting to simplify classification/decision-making in an effort to reduce the need for higher-skilled line staff (i.e., the “dummying down” of community corrections). In this scenario, it is possible to envision a probation or parole agency where line staffs are responsible for case planning and supervision, but other functions (assessments, treatment, and services) are subcontracted to agencies in the private sector.

While improving staff quality and/or privatization are options to consider *even if* you are not using clinical assessments, the evidence available from major reviews of the available research certainly suggests that you will not *improve* the risk assessment system using clinical

assessments, because actuarial risk assessments consistently “outperform” clinical risk assessment procedures (see Harris, this issue; and Gottfredson and Moriarty, this issue). However, we should point out that the new generation of “actuarial” risk assessment instruments currently being used in community corrections agencies—including the popular LSI-R instrument discussed in several articles in this issue—actually requires both objective and subjective (or clinical) assessments by line staff. In fact, the distinction between actuarial and clinical assessment is becoming blurred, with consequences for line community corrections personnel (and management) that are important to consider. It is our contention that if we continue further in this direction, then changes in either staff quality or in the privatization of the assessment function may be needed.

According to a recent review by Brumbaugh and Steffey (2005), three of every four probation and parole agencies in this country employ “objective” risk/needs instruments to classify offenders, using either the shorter Wisconsin risk/needs assessment instrument or the longer, 54-item LSI-R mentioned earlier. Both instruments require line staff to make both *objective* assessments (such as prior convictions, current employment) and *subjective* assessments (such as extent of drug problems, attitude, mental health). Not surprisingly, the results of a number of inter-rater reliability studies reveal that line community corrections staff are much more consistent in their scoring of objective than subjective items (see, e.g. Austin, this issue; Byrne and Robinson, 1991; and Harris, this issue).

The use of a large number of items in a risk instrument is likely to exacerbate the inter-rater reliability problem. Austin (this issue) for example, pointed out that (in an inter-rater reliability study he conducted) of the 54 items included in the LSI-R (37 yes/no items and 17 likert scale items), only 16 items had an agreement rate of 80 percent or higher, with 38 other items scoring in the 60-70 percent range. Overall, Austin found that these scoring differences on individual risk items resulted in disagreement on the scoring of the offender’s risk level in 29 percent of the 120 cases reviewed by the two staff members included in the inter-rater reliability test.

We can only speculate about how such differences in the scoring of individual risk items would affect risk assessment (and the classification of offenders into high-, medium-, and low-risk categories) across an entire department. However, Austin’s research certainly suggests that new strategies need to be developed to improve the level of inter-rater reliability *before* the agency embarks on the time-consuming risk assessment construction/validation process. Our recommendation is to collect data on fewer items, focusing primarily on objective items that are relatively easy to code. Austin (this issue) found that he could improve both the reliability and validity of the LSI-R by focusing on a subset of only 8 of the 54 original LSI-R items. According to Austin, “not only do these items have better predictive ability, but also they reduce the “high-risk” category” (this issue). Since most observers (see, e.g., Lowenkamp and Latessa, 2005) recommend providing the highest level of supervision and services to high-risk offenders, the cost effectiveness of a more precisely defined—and smaller—high-risk classification category should be obvious.

In addition to Austin’s research, the findings from other inter-rater reliability studies indicate that line staff characteristics (such as age, gender, race, location, experience) will likely affect the scoring of risk assessment instrument items in ways that are important to consider. For example, Byrne and Robinson (1992) identified gender bias as a potential problem affecting inter-rater reliability. In their study of inter-rater reliability among 130 probation officers, they distributed two different versions of a “case study”: in version A, the juvenile (Sandy) was described as female; in version B, the juvenile was described as male. There were no other differences between the two “case studies.” Significant differences in overall risk scoring were identified, with the female version of the case receiving *higher* scores than the male version of the same case, resulting in a greater proportion of the female cases being classified as high risk (40.4 percent) than their “male” counterparts (33.1 percent).

We suspect that in addition to variations in scoring and consistency due to *offender* characteristics (such as gender, race, class), there will be variation in scoring and consistency due to the characteristics of the line probation/parole officers completing the assessment (see, e.g. Byrne and Robinson, 1990). The findings from both the Austin study and Byrne and Robinson

study underscore the importance of conducting an inter-rater reliability study, not only to support initial risk instrument development, but also to examine the very real possibility that bias (related to both the characteristics of offenders and the characteristics of line staff) is having a detrimental effect on the risk assessment process.

One other area where clinical (subjective) judgment enters into the classification process is in the agency's risk scoring "override" policy. While there will undoubtedly be circumstances where an offender will be either over-classified or under-classified by line staff members, and/or by management decisions to ignore risk altogether (such as offense exclusions for sex offenders), it is critical that the level, type, and circumstances of over-ride usage be monitored on an ongoing basis. A simple rule of thumb for this type of review is to apply a 10 percent rule: if more than 10 percent of the agency's risk scoring decisions are being changed, then the agency has a problem in this area that needs to be resolved.

Finally, our discussion of the utilization of clinical and actuarial risk instruments in community corrections would be incomplete without mention of the "validation" problem. According to a recent study by Hubbard, Travis, and Latessa (2001), "only 30 percent of the agencies that use an assessment instrument reported that the instrument was validated for their local population" (as summarized by Brumbaugh and Steffey, 2005:59). Without the completion of the necessary validation research, there is no way of knowing whether the risk instrument being used by a particular agency results in an accurate classification of offenders into the low-, medium-, and high-risk categories used to allocate scarce probation and parole resources, both in terms of supervision and in terms of services.

While there certainly has been much discussion of the need to apply the "risk principle" (risk, needs, responsivity) to offenders supervised in the community corrections system (see, e.g. Andrews, et al., 1990, Latessa and Lowenkamp, 2005), it appears that the determination of "risk level" may be inaccurate for a significant number of these offenders. Recent attempts to *improve* the risk assessment process using actuarial instruments may have made matters worse, because of the inter-rater reliability problems associated with the more complex risk assessment instruments currently being used, such as the 54-item LSI-R. The implications of our brief review of the use of actuarial vs. clinical assessment are straightforward: in order to improve both reliability and validity, risk assessment instruments need to be designed using a small number of objective risk items and tested (for reliability and validity) on an ongoing basis.

We anticipate that the continued development of LSI-R type assessment instruments, along with the use of offender-specific assessment devices (for categories such as sex offender, mentally ill offender, substance abusing offender, multiple-problem offender) will require more qualified line staff and/or the use of private sector assessment centers. However, current line staffs in community corrections agencies are certainly qualified to classify the risk level of offenders using the simplified versions of the actuarial risk instruments advocated by Austin (this issue). The challenge for community corrections is to allocate resources to offenders placed in various risk classification levels in a manner that maximizes the system's overall effectiveness. Getting the assessment "right" is the critical initial step, but it must be followed by improvements in treatment classification, and subsequent case planning strategies (see Taxman, this issue).

[back to top](#)

Issue 3: The Need to Combine Individual and Community Risk Assessment

A number of recent evidence-based reviews of the research in community corrections have identified statistically significant, but modest (10 percent) recidivism reduction effects associated with a variety of community treatment strategies (see Welsh and Farrington, 2006). We suspect that the recidivism reduction effects identified in these studies would be even more pronounced if individual-level assessments of risk were combined with community-level risk assessments (Byrne, 2006; Pattavina, Byrne, and Garcia, 2006), based on the premise that community-level risk assessment is a necessary "first step" in the community change process.

We offer this assessment based on two related factors: first, there is a large body of research supporting the notion that an individual's risk of re-offending is affected—both positively and negatively—by the community in which he/she resides while under community supervision (Sampson and Bean, 2005; Sampson and Raudenbush, 2004; Pattavina, Byrne, and Garcia, 2006). Second, the treatment resources available to offenders will also likely vary by the “risk level” of the neighborhood, with higher-risk neighborhoods offering fewer (and lower quality) treatment options to offenders living in these areas (Jacobson, 2006). Accuracy of the *individually-based* risk classification system will likely improve with the inclusion of overall community risk level (high vs. low/medium risk, for example, based on offender density and/or the area's crime rate), along with selected community “risk” characteristics (such as unemployment rate, proportion of residents living in poverty, size/characteristics of first generation immigrant population). Similarly, the accuracy of the individually-based *treatment* classification system (linking offenders at different risk levels to appropriate treatment) would also be improved by an assessment of community risk level, because this classification decision could be based on an assessment of the likely impact of community culture (such as attitudes toward substance use, criminal thinking, etc.) on the attitudes and behavior of offenders residing in “highrisk” and low/medium-risk neighborhoods (Sampson and Bean, 2005).

[back to top](#)

Concluding comments

While there have been significant improvements in the individual offender assessment procedures used by community corrections agencies over the past two decades, our brief review suggests the following: 1) we need to conduct high quality experimental research on the effectiveness of both risk and treatment classification systems, using risk reduction as our primary outcome measure; 2) we need to consider simpler alternatives to both the general (e.g. LSI-R) and offender-specific (e.g. mentally ill, substance abuser, sex offender) risk assessment devices; and 3) we need to incorporate community-level risk factors into our current assessment system.

[back to top](#)

[References](#)

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

[Home](#)

From Theory To Practice: The Lifecycle Document For The Results-Based Management Framework For The Federal Probation And Pretrial Services System

[Endnotes](#)

[The federal probation and pretrial services system is making steady progress towards a comprehensive outcome-measurement system ¹ that will allow the federal judiciary's policymaking body, the Judicial Conference of the United States, and chief probation and pretrial services officers and system managers to make decisions based on ongoing empirical analysis of reliable and pertinent data. We expect to have much of the data collection infrastructure in place in fiscal year 2007 and all of it in place in fiscal year 2008. We will then be able to identify various recidivism-reduction strategies used by different districts and will be able to track large cohorts of offenders to determine what strategies appeared to make a difference under what circumstances for what types of offenders.

Pending completion of their own internal outcome-based system, federal probation and pretrial services officers are eagerly adopting research-supported recidivism-reduction strategies identified in the growing body of research into state and local programs known as Evidence-based Practices. ² When the federal outcome-based system is fully implemented, it will be possible to determine whether those evidence-based practices made a significant difference in measurable outcomes and, if so, whether they should be implemented in other districts.

What follows is a report prepared by Caliber, An ICF Consulting Company under contract with the Administrative Office of the United States Courts. This report provides some useful background and describes precisely how we have built and will continue to build our outcome-based system. Much of the body of the report is reproduced below, except for most of the attachments and section 6, which details nuts-and-bolts implementation.]

1. Project Background

The federal probation and pretrial services system is developing a results-based management framework that will, in the future, allow it to better assess performance—and make programming and resourcing decisions—based on what it accomplishes rather than solely on what it does. The flow chart below shows the steps involved in developing the framework, and highlights where we are in the process.

This focus on results, and the work done to date to define the system's mission, goals and desired outcomes, stem from a number of complementary influences and projects.

1. In 1999, the Administrative Office entered into a contract with a team of independent consultants, led by IBM, to conduct a strategic assessment of the federal probation and pretrial services system. The overarching recommendation from that assessment — presented first to the Administrative Office in 2003—was that the federal probation and pretrial services system become a results-driven organization with a comprehensive

performance measurement system.

2. In 2000, the AO Director appointed an Ad Hoc Supervision Work Group comprised of supervisors, deputies and chiefs from seven districts and a representative of the Federal Judicial Center to update the supervision policy monographs. As part of its work, the group reviewed relevant statutes and mission statements to identify the desired outcomes and goals to be served by the pretrial services and post-conviction supervision functions. These outcomes and goals were incorporated into revised supervision policy documents approved by the Judicial Conference of the United States in 2003.
3. Strategic planning sessions were conducted at the 2000 and 2002 Federal Judicial Center's National Chiefs Conferences. The 2000 conference produced a "Desired Futures" roadmap the first element of which was: "Desired Outcomes are clear, measured and results are communicated." The 2002 conference resulted in a "Charter for Excellence" that sets forth broad system goals and values (see Attachment 1).
4. In September 2003, one of the IBM strategic assessment consultants facilitated a strategic planning session at a meeting of the Chiefs Advisory Group to translate the broad "Charter for Excellence" statements into more specific "Operational Goals."

The operational goals developed by the Chiefs Advisory Group were combined with the desired outcomes set forth in the revised supervision monographs to form the basic structure of the results-based management framework (see [Attachment 2](#)). This concluded the initial goal-setting stage of the framework development process. ³

The current stage of the process is technical: The development of operational definitions and associated measures for each "desired outcome;" and of statistical approaches to analyze the information that will assure "apples-to-apples" comparisons and allow benchmarking with other programs. To assist in this phase:

- In March 2004, the AO director appointed an Ad Hoc Expert Panel on Outcome Measurement Methodology. This panel is comprised of the directors of research for the Federal Bureau of Prisons, the Federal Judicial Center and the District of Columbia's Court Services and Offender Supervision Agency; and academics and Administrative Office staff with extensive backgrounds in criminal justice/substance abuse research and government performance measurement systems. The panel met twice in 2004 and produced recommendations for measuring the concepts of recidivism and substance abuse (see Attachment 3).
- In April 2005, the AO entered into a contract with Caliber Associates to provide additional support in developing, coordinating and presenting recommendations for the technical aspects of the results-based framework.

This final product from the technical phase includes recommendations, to be circulated to system staff and stakeholders for broad system comment, that span the lifecycle of the results-based management framework. The report addresses:

- How to measure a variety of outcomes—including offender compliance, positive change, and crime reduction;
- What data are needed to construct the recommended measures;
- What analytical methodologies can be used to assess how these results are affected by supervision interventions as well as a variety of case, offender and community factors?
- What tasks are necessary to fully implement the framework; and
- How to institutionalize the framework within the federal probation and pretrial services system?

The recommendations represent "state of the art" measurement and analytical approaches that are being used by other performance-based systems, program evaluations and/or academic research in

criminal justice and related areas such as substance abuse.

2. Post-conviction Supervision Logic Model

Building on the results of the goal-setting stage of this project, the next step was to develop a logic model that depicts the underlying assumptions about how “what the system *does*” affects what it is trying to *accomplish*; and what other factors—e.g., characteristics of the offenders to be supervised, the requirements and restrictions of their sentences, and the system resources devoted to carrying out the supervision mission—are expected to influence this relationship. The [logic model for post-conviction supervision](#) is used throughout this document to illustrate the technical concepts that are incorporated into the framework design.

This logic model has been refined twice since its development following the goal setting stage based on input from the Expert Panel and Caliber Associates. It will continue to be a work in progress that evolves to incorporate feedback from system staff and stakeholders, and results from empirical testing of the posited relationships. A similar logic model was developed for [pretrial services](#); which will be incorporated into the framework and follow a similar evolution process.

2.1 Components of the Logic Model

The [post-conviction supervision logic model](#) has six components: inputs, process (activities), process outcomes, intermediate outcomes, ultimate outcomes, and mission. Each component is described below.

Inputs

Inputs are characteristics of the offender population and the working environment that are hypothesized to affect expected outcomes *regardless* of system interventions. For example, prior research indicates that offenders with a lengthy prior record are more likely to become re-involved in criminal activity than those with no or a minimal prior record. This leads to a working assumption that, regardless of supervision interventions, districts that have a high percentage of first offenders will have a lower recidivism rate than those with a low percentage of first offenders.

Inputs are used in the analytical model as “control” variables to account for the effects of factors that explain differences in outcomes across offices, districts and time that are not related to system interventions. They may also be used as stratification categories to display outcomes based on key groupings, e.g., by type of supervision (probation/parole/supervised release) in reports.

The current model includes as inputs those factors identified in the research and program evaluation literature as related to criminal justice goals. These include:

- Offender characteristics (e.g., prior record, employment; family/community connections, demographics);
- Characteristics of the instant offense (e.g., class and category);
- Sentence parameters (e.g., length of prison and supervision terms imposed and served, conditions imposed);
- Office/community characteristics (e.g., location, size, socio-economic indicators);
- Officer characteristics (e.g., experience, demographics, education);
- Supervision resources (e.g., supervision staffing, contract budgets, technological support).

The inputs categories will be further defined and the categories and their specific elements assessed for adequacy by system staff and stakeholders as part of the framework implementation. These inputs will also be applied, as appropriate, to the [pretrial services model](#).

Process (Activities)

Process refers to activities undertaken by the system—practices, programs and interventions—that implement the supervision function. As an example: An officer conducts an initial assessment investigation, identifies lack of stable employment as a risk, and refers the offender for job counseling or to a job referral agency. In the analytical model, the process variables define “what we do” for purposes of assessing the basic relationship of how “what we do,” relates to what we are trying to accomplish.

The current logic model includes only the most general process categories, e.g., investigation, assessment, monitoring, referral, and assistance. Detailed input on the specific processes that should be included in the model will be sought from system staff and stakeholders—the experts in identifying and defining salient system activities—as part of the initial next step in implementing the framework.

Process Outcomes

Process outcomes describe *offender* actions that occur as a result of system activities. For example, in response to an employment referral, the offender registers with an employment service or completes “x” hours of employment counseling.

Process outcomes enter the analytical model as both an outcome of the service delivery process and as an input (control) for assessing interim and ultimate outcomes. For example, “number of hours of employment counseling” is a measure of how successful an officer’s employment referrals are in engaging offenders in employment services. This measure is also a “control” when addressing the extent to which an interim outcome, such as improved employment, might be attributed to the supervision intervention.

The model includes only broad categories of compliance with each of the four major types of conditions: Restrictions, correctional programming, service, and financial. These will be expanded as a result of comments received from system staff and stakeholders.

Intermediate Outcomes

Intermediate outcomes are changes in offender behavior that are themselves desirable and believed also to be precursors of the ultimate outcomes. Specifically, the intermediate outcomes in the model are defined as a desirable *change* in a circumstance that has an *empirically proven relationship* to successful supervision and that is within an officer’s *authority and sphere of influence*.

Three intermediate outcomes are included in the current model: Reductions in substance abuse, improvements in employment, and improvements in other life skills. Each of these changes is believed to relate to the likelihood of criminality during the period of supervision and beyond. Each is also expected to relate to certain “sentence execution” outcomes (e.g., improved employment, enhanced earning capacity, more money to pay restitution).

The list of intermediate outcomes currently in the model does not include all of the operational goals that emerged from the Chiefs Advisory Group’s strategic planning session in 2003. This results from the limitations that were subsequently attached to the definition of intermediate outcomes. For example, “Improvements in Mental Health” is *not* listed as an intermediate outcome because there is no empirical association between general mental health problems and criminality. This does not mean that mental health issues will not be considered, but rather that their relationship to ultimate outcomes will be considered in terms of offender and sentence characteristics (e.g., mental health needs and conditions as inputs), referrals for mental health counseling (process), and the offender’s participation in that programming (process outcomes).

Ultimate Outcomes and Mission

The ultimate outcomes are set forth in *The Supervision of Federal Offenders, Monograph 109*, which establishes Judicial Conference policies related to post-conviction supervision. These outcomes are: To execute the sentence and to protect the community during the period of

supervision and beyond.

These outcomes stem directly from the system mission endorsed by the Judicial Conference in September 1993—To protect the public and to assist in the fair administration of justice—supplemented by the statutory provisions that establish the duties of probation officers and the purposes that community sentences are to serve.

Two of the ultimate outcomes—minimizing criminal activity during the period of supervision and beyond—relate to the system mission to protect the public. The other ultimate outcomes measure the impact of compliance with release conditions (e.g., restoration of victims). These serve as direct measures of sentence execution and surrogates for the mission to assist in the fair administration of justice—with no expectation that they will affect the system’s public protection mission. Examples from the literature of more precise operational measures for process outcomes, intermediate outcomes, and ultimate outcomes are provided in Section 3 and further delineated in Attachment 5: Key Element Definitions and Attachment 6: Data Matrix.

2.2 Relationships among Components

The arrows in the logic model indicate the specific expected relationships between components that the analytical model will be designed to test. As described in Section 4, statistical techniques will be applied to test the relationships depicted.

The analysis will test a complete thread of the model, starting from left to right. Basic and advanced techniques will be used to test both direct and indirect and unidirectional and bidirectional relationships, while controlling for inputs that are primarily static and outside the control of the officer. The results will move the system beyond a description of the offender population and individual outcomes to a more complex assessment of the “theory of change” and the interconnectedness of process and outcomes for post-conviction supervision. Similar relationships will be tested for pretrial services.

3. Operationalizing Post-conviction Supervision Outcomes

This section further defines the process, intermediate, and ultimate outcomes in measurable terms. In order to empirically test the hypothesized relationships between post-conviction processes (activities) and outcomes of the offender population, it is necessary to first identify appropriate measures for each outcome. A data matrix was developed based on a review of current evaluation research. ⁴The matrix organizes each outcome into the following categories:

- Type—The concept of interest.
- Definition—A brief description of each concept. The descriptions are important to ensure standardization in how data are defined across districts.
- How Operationalized—How the concept will be measured.
- Data Element—The specific piece of data to be captured so that certain tests can be performed to help answer questions of interest. Data may be reported in days, weeks, dollars, cents, or by selecting “yes” or “no” options or other response categories.
- Level of Measurement—The level at which the data will be measured (nominal, interval, ratio, ordinal). This is important for determining the type of analysis that can be performed using each measure.
- Data Source—Where the identified data may be obtained. In some cases, the data may not currently be collected and therefore, the data source will need to be determined by system staff and stakeholders.

3.1 Process Outcomes

A process outcome represents the immediate outcome for the offender as a result of system activities. The four process outcomes are: compliance with restrictive conditions, participation in correctional programming, compliance with service conditions and compliance with financial conditions. Examples of the types of data that could be collected for each outcome are described below.

- Comply with restrictive conditions (e.g., home confinement conditions, halfway house placement, employment conditions, prohibition of contact with victim/minor/associates, remote location monitoring, and nighttime/weekend jail requirements). The process outcome measures for all conditions will include: a dichotomous measure of compliance (complied or did not comply) and number of noncompliant events involving condition. An additional measure for any condition with an associated time component will be calculated based on days imposed vs. days completed. The process outcomes will be calculated separately for each restrictive condition and for the restrictive category as a whole.
- Participate in correctional programming—Measures of this process outcome will include: number of days from start to end of program (duration); number of hours of service (per week and total); number of sessions attended vs. sessions scheduled; end-of-treatment provider assessment of quality of participation (5-point scale); and completion status (successful/unsuccessful completion). The results will be presented by type of program (substance abuse, education/employment/job training, mental health treatment, sex offender, life skills, basics), funding source (no cost, contracted, other government program, or self-insured), and modality of treatment (inpatient, individual, group, family-individual, or family-group).
- Comply with service conditions—Service conditions consist of a requirement for the offender to complete hours of non-paid community service. Compliance with service conditions will be determined by number of community service hours completed vs. the number of hours imposed. A dichotomous compliance status measure (complied or did not comply) can then be calculated based on whether the offender successfully completed the imposed hours of non-paid community service.
- Comply with financial conditions—(e.g., fine, restitution, special assessment, no new debt/credit, cooperate with IRS; child support enforcement) Measures for all conditions will include: a dichotomous measure of compliance (complied or did not comply) and number of noncompliant events involving condition. An additional measure for any financial condition with an associated amount will be: amount expected by payment schedule vs. amount paid. The process outcomes will be calculated separately for each financial condition and for the financial category as a whole.

The process outcomes data described above will be used in the analysis described in Section 4 as both dependent variables (predicted outcome of system activities) and independent variables (predictors of intermediate outcomes).

3.2 Intermediate Outcomes

An intermediate outcome represents the expected immediate result of the process outcomes. Examples of the data to be collected for the three intermediate outcomes depicted in the [post-conviction supervision logic model](#)—reduce substance abuse, improve employment, and improve other life skills—are described in greater detail below.

- Reduce substance abuse—Measures of substance use *during the period of supervision* will include drug test results, self-admissions, and substance-related re-arrests. These measures will be used to create a dichotomous measure of substance use (used or did not use). The outcomes will be presented by time of event (before treatment, during treatment and after treatment) and type of substance. The analysis will assess “change” by comparing substance use during supervision with the offender’s status at the start of supervision based on such factors as: prior diagnosis of addiction/abuse (Y/N); evidence

of use at time of instant offense based on admission, positive pretrial/presentence drug test, offense involving drug/alcohol use? (Y/N); drug(s) of choice; and number of prior treatment experiences.

- Improve employment—Measures of employment may include: change in employment status (unemployed, unemployed but seeking employment, part-time employment, full-time employment), length of employment (calculated as the percent of time offender was employed during period of supervision), and amount of wages.
- Improve other life skills—There is still a question as to the type of life skills that should be identified as outcomes in the logic model. Specifically, which life skills meet the intermediate outcome criteria: “A desirable change in a circumstance that has an empirically proven relationship to successful supervision and that is within an officer’s authority and sphere of influence.” Potential straightforward areas are level of educational and new vocational/advocational skills. Can or should the area be expanded to encompass such topics as family stability and community stability without overstepping appropriate bounds on officer authority?

As with process outcomes, the above intermediate outcomes will be tested as dependent variables (the result of participation in correctional programming) and independent variables (predictors of future criminal activity and victim restoration). As shown in the logic model, the relationship (as either independent or dependent variables) between intermediate outcomes also will be tested. Additionally, the intermediate outcomes will be tested as mediating variables between process outcomes (participation in correctional programming) and ultimate outcomes (minimized criminal activity).

3.3 Ultimate Outcomes

An ultimate outcome is the long-term result of the system activities for the offender. The ultimate outcomes also reflect achievement of the mission of the federal probation and pretrial services system. The five ultimate outcomes that best reflect the mission include: minimize criminal activity during the period of supervision, minimize criminal activity beyond the period of supervision, maximize victim restoration, defray costs to the government, and maximize compliance with release conditions. The analysis of data on these ultimate outcomes will help system staff and stakeholders better assess if the missions of protecting the public and assisting in the fair administration of justice are being achieved. Each ultimate outcome is discussed below.

- Minimize criminal activity during the period of supervision—The primary measure of criminal activity during the period of supervision is whether an offender was arrested for a new offense. Technical violations are not counted as a new offense. The analysis will also examine the time to arrest (length of time before the arrest for a new offense). The results will be presented overall and by offense type (e.g., violent, property, drug, public order, weapon, immigration) and offense level (felony, misdemeanor, petty).
- Minimize criminal activity beyond the period of supervision—Similar measures for criminal activity beyond the period of supervision will be reported as described above.
- Maximize community restoration—Community restoration will be measured as the amount of payments made to victim special assessments and, where required, the Victims’ Crime Fund. In addition, community restoration will be measured by the amount of fines paid, etc.
- Maximize compliance with release conditions—Compliance with release conditions will include: a dichotomous measure of any noncompliance (Y/N), number of instances of noncompliance, time to first noncompliance (months), and time free of noncompliance at inactivation/termination of supervision (months).

Ultimate outcome data enable system staff and stakeholders to test whether the system activities (processes) are leading to the long-term outcomes that the federal probation and pretrial services system is tasked with achieving. Furthermore, these data will allow system staff and stakeholders to assess how well they are doing at meeting their mission to protect the public and fairly administer justice.

4. Data Analysis Plan

The data analysis plan describes a recommended approach to testing the relationships depicted in the logic model. That is, the plan presents the statistical techniques, progressing sequentially from simpler to more sophisticated levels of analysis, that will provide system staff and stakeholders with both a description of the offender population and outcomes and a more complex assessment of the hypothesized “theory of change” and interconnectedness of process and outcomes (e.g., direct, indirect, unidirectional and bidirectional relationships) depicted in the [post-conviction supervision logic model](#). Specifically, the plan describes basic and advanced statistical techniques that can be applied to test these relationships, while controlling for inputs that are primarily static and outside the control of the probation officer. The analysis plan is organized into three stages: data quality, data reduction, and data analysis. Each is presented below.

4.1 Data Quality

Before any data analysis is conducted, all data will need to undergo standard checks for quality to make sure there are no data entry or transmission errors. Checking for data quality is typically a two-step process that involves detection and then correction of errors in a data set. Cleaning and preparing data is an often neglected but extremely important step in the analysis process. The saying “garbage-in-garbage-out” is particularly applicable where large data sets collected via some automatic methods (e.g., via National PACTS Reporting, National Crime Information Center (NCIC), etc.) serve as the primary input into the analysis. The most common sources of error include data entry errors, such as typing errors, column shift (data for one column being entered under the adjacent column), which often results in invalid responses, general coding errors, which may occur during data collection or entry and may be difficult to detect unless you look for outliers or unusual relationships between variables, and not recoding missing data, which can result in inflated mean scores and the like. The first step to ensuring quality data begins with ensuring clear understanding and procedures for collecting and entering data and systematic review before data are submitted or made available for extraction. Once data are submitted, the next step in the data quality assurance process is to detect and clean the data for errors.

Error Detection

There are three common procedures for detecting data errors that should be followed. These include:

- Review descriptive statistics. Using software, such as SPSS, the following review of descriptive statistics can identify data errors:
 - Look at minimum and maximum values to determine if data fall outside the acceptable range.
 - Look for presence of 0’s and 999’s (or 9999, etc.) shown in frequency tables, graphs or histograms to indicate missing values.
 - Look at means, medians, and standard deviations. For example, if the median differs much from the mean value, it is important to investigate the overall distribution of values for outliers.
 - Assess frequencies. By examining frequencies, it is possible to detect unequal distribution in categories such as age, sex, and race that are outside what would normally be expected for a particular population.
- Conduct logic checks. Errors in data can be detected simply by determining whether the

responses seem logical. For example, you would expect to see 100% of responses, not 110%.

- Examine bivariate outliers. Some data errors only appear when two variables are compared. To detect such error, it is important to look for outliers, or values of a variable that are far different from the expected values. These errors can be detected by examining bivariate associations and scatter-plot graphs to check for deviations in expected associations between variables.

Once the data errors are detected, there are several techniques that should be followed for correcting the errors. These are described below.

Error Correction

Once errors are detected, it is important to know how to handle them appropriately so the data can be analyzed without losing their integrity or robustness. There are slightly different ways to deal with error in independent (or predictor and control) variables and dependent (outcome) variables.

Independent Variables

When there are a minimal number of errors, the values are generally recoded to “missing.” What this means is that the suspicious values are counted as missing data since they are not within an acceptable range. If there are many error values, then it is important to check to see if some of the values of the outcomes are the same for missing and nonmissing values for the independent variables. If so, then there is less chance of bias in the analysis. If not, then it is possible that the data is not good and that the variable should be discarded or used with caution. Various imputation-based procedures to fill in missing values (series mean, mean or median of nearby points, linear interpolation, linear trend at point) will need to be considered. Other more complex imputation-based procedures (regression imputation, non-ignorable missing-data models, Heckman’s two-step statistical process) may also need to be used. [5](#)

Dependent variables

If there are few data errors, values can be set to “missing” using one of the imputation-based procedures determined to be most appropriate. However, it is important to use caution when setting many values to “missing,” especially if multiple variable analysis will be conducted. It may be necessary to set the error values for the outcome or independent variable to the data set mean or the group mean (maybe by age, type of offender, etc.). This should result in a histogram with a more normal distribution of values. Once quality of the data has been checked and the necessary steps taken to correct for problems or errors, data analysis can proceed from basic to more advanced techniques described below.

4.2 Data Reduction

Data reduction is a process often applied where the goal is to aggregate or amalgamate information contained in large data sets into more manageable and reliable information. Data reduction techniques can include simple tabulations, aggregations, or more sophisticated techniques, such as clustering, principal components analysis, and path analysis. Each of these recommended data reduction methods is described below.

Aggregation

Aggregating or transformations of data are techniques often used to reduce or optimize your data. Data can be aggregated, for example, by subgroups to move from individual case records for thousands of individual offenders to mean scores for subgroups of offender populations based on certain criteria (e.g., criminal history, district, gender, etc.). Additionally data can be transformed by creating dichotomous variables (presence/absence of new offense) from continuous variables (number of new offenses) or composite scores or constructs from multiple measures (risk

assessment score). When transforming data, especially creating composite measures or constructs, it is important to use other techniques, such as those described below, to determine which variables should be combined to create a new variable.

Cluster Analysis

Cluster analysis is a multivariate analysis technique that seeks to organize information about variables so that relatively homogeneous or similar groups, or “clusters,” can be formed. To use this technique, it is important that the clusters formed be highly internally homogenous (members are similar to one another) and highly externally heterogeneous (members are *not* like members of other clusters). Cluster analysis can be used to combine “similarity” measures as well as measures that are proxies or associations. However, it is first necessary to standardize your data since clustering often involves combining items measured on different scales.

Principal Component Analysis

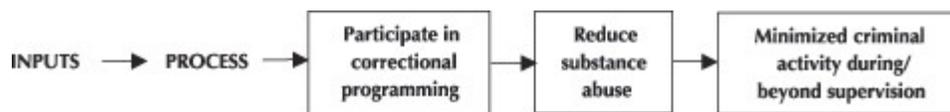
Principal component analysis is used to reduce the number of variables or factors for inclusion in your analysis. Specifically, this method of analysis is used to combine two or more *correlated* variables into a single factor. Principal component analysis helps reduce redundancy in your measures by identifying and combining those that are highly correlated into a new variable, while minimizing the variance around the new variable. ⁶

Path Analysis

Once you have arrived at a set of measures that represent the variables of interest (inputs, processes, process outcomes, intermediate outcomes, and ultimate outcomes), it is necessary to determine whether the relationships or paths among those variables presented in the [post-conviction supervision logic model](#) can be supported by the data. That is, do the data fit the model? Path analysis calculates a path coefficient, which shows the direct effect of an independent variable on a dependent variable in the path model. This information is then used to calculate a goodness-of-fit statistic. The statistic determines how well the data fit the model, or stated another way, the statistics help identify the best fitting models for the data. ⁷ For example, based on the analysis, it may be determined that certain predicted or hypothesized relationships shown in the logic model are not supported by the data. Or, the analysis may uncover other relationships (direct or indirect) that are not currently represented in the model. It is important to test these relationships using path analysis or other more sophisticated techniques, such as structural equation modeling (a variation of path analysis involving multiple indicators of variables in the model), to ensure the “big picture” is accurate before attempting to further describe your data and examining more minute relationships among variables (see basic and advanced analysis sections below). ⁸ Based on the results, it may be necessary to modify the model (add and/or remove arrows depicting relationships, add/or remove variables, etc.).

4.3 Data Analysis

The data analysis stage is presented in two parts: basic analysis and advanced analysis Each is presented below. To help illustrate the rationale behind each recommended analytic technique, examples drawn from the substance abuse “thread” of the logic model shown below are discussed.



Basic Analysis

Once the steps necessary to reduce or optimize the available data have been completed, the first or basic level of analysis to be conducted of the input, process, and outcome data will use descriptive statistics. Descriptive statistics should be used to:

- Describe the basic features of the data
- Provide simple quantitative summaries about the measures
- Provide the basis for subsequent data analysis.

In general, descriptive statistics will describe “what is,” or what the data show with respect to a given variable.

Univariate Analysis

Univariate analysis of the data is used to examine across cases or groupings (e.g., offender cohorts, offender type, districts, etc.) one variable (or outcome) at a time. The major characteristics or descriptive statistics to be examined for each variable include:

- Distribution (frequency distribution, percentages)
- Central tendency (mean, median, mode)
- Dispersion (e.g., range, variance, standard deviation).

These descriptive statistics will provide system staff and stakeholders with basic information about the offender population as a whole. [9](#)

Testing the “thread” of the logic model shown above, the following sample questions can be answered using univariate analysis:

- What are the demographics (e.g., distribution by race, age, gender, risk level, seriousness of previous criminal offense(s)) of the offender population entering supervision with an identified substance abuse problem?
- What percentage of offenders with an identified substance abuse problem participate in substance abuse treatment? What is the average dosage of substance abuse treatment received across this offender population? What is the most common modality of treatment provided?
- What percentage of this offender population abstains from using illegal or other restricted substances during the period of supervision?
- For those offenders with a history of substance abuse who are identified as using substances during supervision, what are the most common drugs of choice identified?
- What percentage of this offender population commits a new offense during supervision? Following release from supervision?
- What is the average length of time to a new offense for this population?

Additionally, by plotting the data using scatter-plots, histograms, bar graphs, or other options, it will be possible to identify potential differences or variation in outcome responses for subgroups of the offender population for further analysis. For example, are there visible differences in the distribution of offenders with a history of substance abuse with a new arrest while under supervision? Are there differences in the distribution of this offender population by reasons for ending supervision (e.g., successfully completed on time, early release, revocation, absconded)? Is there wide variation in the amount of treatment hours provided for this population of offenders?

Bivariate Analysis

The next step in the analysis is to begin testing the significance of the covariates (inputs, process (activities), process outcomes) on the intermediate (e.g., substance use, employment, life skills) and ultimate outcomes (criminal activity, victim restoration, compliance with release conditions) using bivariate analysis techniques, such as correlations, cross tabulations with chi-square statistics, t-tests, and analysis of variance. These techniques are recommended because they are

relatively easy to conduct and provide straightforward interpretations. ¹⁰ The bivariate analysis will provide preliminary indication of the relationship between independent and dependent variables to be further tested with the inclusion of control variables (e.g., offender characteristics, offense characteristics, sentence characteristics, etc.) with more advanced analysis.

For example, analysis of variance (ANOVA) is used to test differences between two or more means. ANOVA helps to determine which variables have a significant influence on an outcome, and/or how much of the variability in the outcome or dependent variable is attributable to each factor. The two types of ANOVAs that can be used to test such differences include:

1 One-Way ANOVA. A one-way analysis of variance is used when the data are divided into groups according to only one factor (e.g., level of risk, type of correctional programming, compliance status). The questions of interest are usually: Is there a significant difference between the groups? and if so, which groups are significantly different from which others? Statistical tests are provided to compare group means, group medians, and group standard deviations. ¹¹

- Multifactor ANOVA (MANOVA). When more than one independent or control variable is present and the factors are crossed (e.g., instant offense by prior record), a multifactor ANOVA is appropriate. Both main effects and interactions between the factors may be estimated.

Using the substance abuse example, a one-way ANOVA might be conducted to test differences in mean number of substance abuse relapses between those offenders with a history of substance abuse who participate in individual substance abuse treatment and those who participate in other modalities of treatment (group, family, and other). ¹² A detailed example and interpretation of the results from an ANOVA are presented in Attachment 7.

Another useful bivariate analysis to be considered is the negative binomial regression model. This technique can be used to test whether different subgroups of offenders commit more frequent acts of noncompliance with conditions of supervision, relapse, and/or new offenses or other outcomes than other subgroups. Negative binomial regression models were developed specifically for the kind of distribution of failures that are likely to be observed with these offender data (i.e., a large portion of the offenders will not fail at all during the time observed, some will fail once, fewer will fail twice, and a handful will fail more often). This type of skewed distribution (if present) would violate the normality assumptions of ANOVA. ¹³

Advanced Analysis

Based on the results of the univariate and bivariate analysis, more advanced statistical techniques (logistic regression, survival or Cox regression) are recommended to test the relationship between process outcomes, intermediate outcomes, and ultimate outcomes after controlling for inputs (e.g., characteristics, sentence, and resources) and process (system activities). The recommended advanced statistical techniques are described below.

Logistic Regression

Logistic regression allows one to predict a discrete outcome, such as group membership, from a set of variables that may be measured at any level (interval, ratio, ordinal, or nominal) or a mix of levels. Generally, the dependent or outcome variable is dichotomous, such as presence/absence or compliance/noncompliance. Logistic regression calculates the probability of success over the probability of failure, presenting the results in the form of an odds ratio. For example, logistic regression can tell us the probability that an offender will reoffend after controlling for various input or process factors. Logistic regression also provides knowledge of the relationships and strengths among the variables.

For example, it is possible to use logistic regression to test the relationship depicted in the logic model between reduced substance abuse and minimized criminal activity beyond the period of supervision. The independent variable would represent a measure of substance abuse during supervision. This could include a dichotomous variable representing presence or absence of

relapse created from existing measures of substance abuse or a continuous variable representing the number of relapses during supervision. The dependent variable would represent criminal activity following supervision. For logistic regression, this would be represented by a dichotomous variable, such as presence or absence of a new offense, substance abuse-related or other type of crime, or other dichotomous measures of ultimate outcomes presented in the logic model. Additionally, it is important to include measures of inputs as control variables into the regression equation, such as characteristics of the offender, offense, and sentence. For ease of interpretation and to best understand the variance accounted for by the control variables, it is recommended that the inputs or controls be entered first in the equation as a block or set of control variables.

Unlike linear regression, the interpretation of the coefficient for logistic regression ($\text{Exp}(B)$) is more straightforward, representing the likelihood of an event occurring. Using the substance abuse example, the results of logistic regression can tell you how much more likely an offender who relapses during treatment is of committing a new (non-substance abuse-related) offense during and beyond supervision than an offender who does not relapse during treatment. With logistic regression, it is also possible to control for a block or set of variables (inputs, process, etc.) in the analysis.

Survival Modeling

Survival modeling is recommended when you want to examine the relationship among independent variables or covariates and the time to events of interest, for example, time to employment, time to relapse, time to recidivism, time to completion of paying restitution, etc. Static models alone are insufficient in this situation because they assume that a rearrest or other outcome is the same regardless of whether it occurred on the first or last day of the period of interest. The timing of these events, however, is a particularly important distinction when considering the public policy and safety implications of supervision. Survival modeling or analysis is an effective statistical technique to use when you want to examine the impact of time-varying covariates on these events. It is a particularly useful technique when comparing groups with varying follow-up periods. Survival analysis handles time at risk by subdividing the follow-up period into smaller observation points. At each of these points, the proportion of the sample that is at-risk for reoffending, for example, is used to estimate the probability of surviving beyond that point. This method ensures that only the characteristics of the population still at risk are used to estimate the time until failure, thereby providing a more accurate prediction of failure. Additionally, survival functions use maximum likelihood techniques that can differentiate between censored and uncensored cases. That is, survival modeling accounts for those cases who survive throughout the follow-up period. ¹⁴

An example of the results of a hypothetical comparison (based on “dummy” data) of the risk of recidivism among offenders with a history of substance abuse who participated in individual treatment compared to offenders who participated in group treatment, after controlling for race, age, and risk assessment is shown in Attachment 7. Like logistic regression, the results of survival modeling are interpreted as the likelihood of an event occurring.

Linear and Multiple Regression

While logistic regression and survival modeling are valuable statistical techniques to use when the outcome or dependent variable of interest is a dichotomous variable, other regressions should be used when the outcome of interest is a continuous variable. For example, if you want to determine whether there is a relationship between offenders who demonstrate a reduction in substance abuse during supervision and the amount of restitution an offender is able to pay, linear regression should be used. If you want to examine the relationship of more than one variable, for example substance abuse, employment, and life skills, on an outcome (amount of restitution paid or number of new offenses), you should use multiple regression. These are all statistical techniques that will enable system staff and stakeholders to further explore threads of the [post-conviction supervision logic model](#) and identify predictors of success.

Trend Analysis

Trend analysis can be used to examine changes in outcomes over time for a given population, as well as to compare trends in outcomes across subgroups. Trend analysis is often depicted by a graph. [This graph](#) depicts hypothetical trends in the behavior of offenders entering post-conviction supervision with an identified substance abuse problem, specifically compliance with restrictive and financial conditions (represents percentage in compliance) over a 12-month period. The graph shows that a greater percentage of offenders were in compliance with restrictive compared to financial conditions during the 12-month period. However, both represent a gradual increase in the percentage of offenders in compliance from January to December. To predict future values for a variable, for example compliance with restrictive conditions beyond December, time series analysis is required. ¹⁵

Analysis Assumptions

It is important to recognize that there are two key assumptions underlying the above analysis plan. The first assumption is that all independent and control variables are treated as exogenous variables. That is, a variable whose variability is assumed to be determined by causes outside the model under consideration. No attempt is made to explain the variability of an exogenous variable or its relations with other exogenous variables. Stated another way, none of the independent or control variables are said to affect (or to be affected by) any of the other variables. It is, however, recognized that these variables may be correlated with one another. The second assumption is that the logic model (multistage model) or segments of the model (single-stage model) being tested are well specified. Under this assumption, for regression analysis, it is possible to interpret the regression coefficient as the expected change in the dependent variable associated with a unit change in the variable in question, while partialing out the influence of the other variables (independent and/or control). ¹⁶

As the logic model and hypothesized relationships are further developed and tested, and data collection refined, additional analysis (e.g., differences of proportions, interrupted time series) will be considered for future assessments.

5. Reporting

The results-based management framework will generate an informative annual research report that effectively shows and explains changes in outcomes over time and reasons or predictors of those changes. While the specific layout of the information in the report will be driven by the type of analysis and research questions to be addressed, a general template for the report is presented in Attachment 8 [omitted here]. Each section of the report is described in detail below.

I. Overview. This section will contain a standard description of the results-based management framework, including the overall purpose and a description of the logic model for post-conviction supervision underlying the framework.

II. Description of methodology. This section will contain a detailed description of the methodology behind the framework. In particular, it will include the research questions to be addressed, data sources and measures examined, selection criteria and resulting sample, and limitations to sample (e.g., limitations to generalizability of results to entire offender population).

III. Data quality and reduction process and results. This section will include a description of the data quality process (see Subsection 4.1) and the results of the analysis, including percentages for missing data, data errors, and other exclusionary factors that result in a reduction of the data set; description of data reduction methods (see Subsection 4.2) and results; and a discussion of any implications or limitations to the analysis as a result of the data quality assessment and data reduction process.

IV. Model fit. Because of the importance of the logic model to the integrity of the results, it is important to present the model fit results (test of the model as a whole) separate from the results for specific relationships depicted in the model. This section will include a description of the

analysis conducted (e.g., principal component, path analysis, structural equation modeling), results (goodness of fit statistics), and an interpretation of the results. The interpretation is important for guiding the subsequent analysis and results. For the initial annual assessment, a complete test of the model may not be possible due to unavailable data.

V. Post-conviction supervision results. This section is critical to understanding what aspects of post-conviction supervision are producing desired outcomes and which areas need modifications or improvements. This section will be divided into two parts: *demographics* and *outcomes*. The *demographics* subsection should begin with the results of the univariate or descriptive analysis in order to provide the reader with a profile or profiles of the offender population, the system itself (districts, offices, regions, etc.), and other contextual factors that are important for interpreting the results. Additionally, it is important to point out that these measures will be used as control variables in the advanced analysis of outcome measures. The results of the bivariate analysis should be presented next to identify and provide evidence for testing specific relationships among processes and outcomes.

The next subsection will present the *outcomes* for post-conviction supervision. The logistic regression should be presented first with a description of the analytic technique (see Subsection 4.3), followed by an interpretation of the results for each run. This should include a description of the control variables, the relationship(s) being tested, and the findings. Where appropriate, graphics should be used to present the findings. Next, the results of the linear and multiple regressions should be presented. Together, the findings from the logistic and linear/multiple regressions will provide important information regarding the specific relationships depicted in the logic model (e.g., does X lead to Y when controlling for A and B?).

The regression results will be followed by a presentation of the survival modeling. A description of the analytic technique (see Subsection 4.3) will be necessary to ensure the reader understands why survival modeling is being used. In particular, it is important to explain that survival modeling is needed to compare outcomes for the different entering offender cohorts compared annually. Additionally, the results of the survival modeling can help predict likelihood of success (or failure) for an offender in the absence of complete follow-on data for all offenders. This will be important for the first several years of the assessments. Again, the use of graphs and charts to present findings is recommended. The results and an interpretation of each run will be provided.

Finally, the results of the trend analysis will be presented. This information will provide the reader with a description of how outcomes have changed over time and for which offender populations. Additionally, the effectiveness of different treatments or interventions can be compared over time. Using line graphs is the most effective method for presenting trend data. An interpretation of the findings will be presented.

VI. Implications and Recommendations for Policy and Practice. This is the most important section of the report. It will begin with a summary of key findings and a discussion of any unexpected findings and limitations of the data. Next, implications of the findings for setting priorities and making policy, programming, and resourcing decisions need to be presented followed by specific recommendations *supported by the results*. If possible, suggestions for implementing the recommendations should also be included.

Subsequent reports will follow the same template but will need to address, if appropriate, the following:

- Changes to data collection (process, sources, measures) and explanation for changes
- Changes to logic model and explanation for changes
- Changes to research questions (or focus of analysis) and reasons for changes.

It is important to include the core sections in each annual report but recognize the content may change and the format may need to be flexible. For example, once pretrial information is incorporated into the framework, it may be necessary to create a separate report template to present these findings or make modifications to the existing template in order to combine the

results into a single report. Feedback should be obtained from end-users and changes made to the report template as appropriate to ensure clear communication of results and the usability of information.

* * *

7. Institutionalizing the Framework

Once the framework has been implemented, it is important to ensure it maintains momentum and continues to receive attention by management. Success can be measured by the extent to which the results-based management framework becomes institutionalized within OPPS. Just as the 2004 Strategic Assessment of the Federal Probation and Pretrial Services System recommended the need to organize, staff, and resource to promote mission-critical outcomes, the same can be said for the sustainability of the framework needed to assess progress toward those outcomes. Accomplishing this task will require adherence to a predefined yet flexible process, strong management support, and a commitment of resources. Each of these critical factors is described in the subsections below.

7.1 Predefined Process

To ensure full implementation of the results-based management framework, a continuous process of assessment, review, and modification is necessary. The specific steps in the process are outlined below.

Ongoing Consistent Assessments

Ongoing assessment is defined as annual extraction, analysis, and reporting on processes and outcomes for offender cohorts entering post-conviction supervision during specific fiscal years. Over time, the plan is to analyze data for three consecutive entering cohorts at a time. It is important to adhere to the criteria for selection each year in order to provide comparisons over time and to identify trends in outcomes. It is critical to ensure comparison of “apples to apples” each year and over time. Additionally, the timing of the extraction and reporting must be consistent each year... To ensure ongoing assessment and the production of results that can be used by managers to make important decisions will require a review of existing personnel to identify those individuals with the necessary skill-set to conduct the assessments (e.g., knowledge of PACTS, ability to apply advanced statistical techniques and interpret results as identified in the analysis plan, etc.).

As baseline data become available for the various components of the logic model, it will be possible to set performance benchmarks against the baseline measures. While baseline measures are indicators of where the system is, benchmarks identify where the system needs to be in the future. It is important to use data (baseline measures) and actual experience when setting benchmarks to ensure they are realistic. This process should be a collaborative effort involving input from the field. Once benchmarks are set, they should be reexamined at least every three years to assess progress and determine if modifications are needed based on the results.

Review and Modify (Feedback loop)

The success of the results-based management framework relies on quality data, appropriate analysis and interpretation of results, and the utilization of the results. In particular, using the results to review and modify the framework is important to the longevity of the model.

As policies and practices change, it may be necessary for the model to change. As with the benchmarks, it is important to review the framework design, process, and results at least every three years to identify any needed changes or modifications to the logic model, measures, data systems, selection criteria, etc. Additionally, information needs of management and the field may shift, requiring changes to the framework. This feedback loop will help ensure a results-based management framework that is responsive to changes over time. Any modifications to the system need to be vetted through key stakeholders, including the field.

7.2 Management Support

It is critical for the results-based management framework to be owned by a specific organizational unit within OPPS. OPPS may wish to consider restructuring the existing organizational structure to create a new unit focused exclusively on results management and the implementation and sustainability of the framework. Whatever approach is taken, it is important that there be a manager whose primary responsibility is the oversight of the framework. The manager must also have the authority or access to the appropriate lines of authority to ensure support of the framework and consideration of recommendations to decisions regarding policy, programming, and resourcing. To the extent possible, it is also preferable that the unit is viewed as independent of the other divisions and branches. This is important to ensure objectivity, neutrality, and ensure the unbiased reporting of findings and recommendations. Additionally, sustainability of the framework requires designated staff with the expertise necessary to ensure data quality, analyze data, and translate results into practical information. Staff also need to be able to make modifications to the framework, including revising the logic model and identifying and operationalizing new measures.

7.3 Commitment of Resources

As with any new effort, it requires resources to get a process up and running and to continue operating over time. A thorough assessment of the resource needs for completing the remaining implementation tasks and sustaining the framework by carrying out the lifecycle plan needs to be conducted. This assessment should be reviewed annually, especially within the first three years as changes and modifications requiring additional resources are likely. It is clear that staff and resources need to be organized to support the framework.

8. Next Steps

This lifecycle document is intended to serve as the primary document that describes the content of the results-based management framework, the analytic approach to the data, implementation process, and plans for institutionalizing the framework within OPPS. It is important that the information presented in this document be reviewed by system staff and stakeholders in order to verify the information, fill in gaps, review recommendations, and resolve unanswered questions. Once the framework is fully implemented, it will provide system staff and stakeholders with the information needed to better assess performance—and make programming and resourcing decisions—based on what the federal probation and pretrial services system accomplishes rather than solely on what it does.

[back to top](#)

[Endnotes](#)

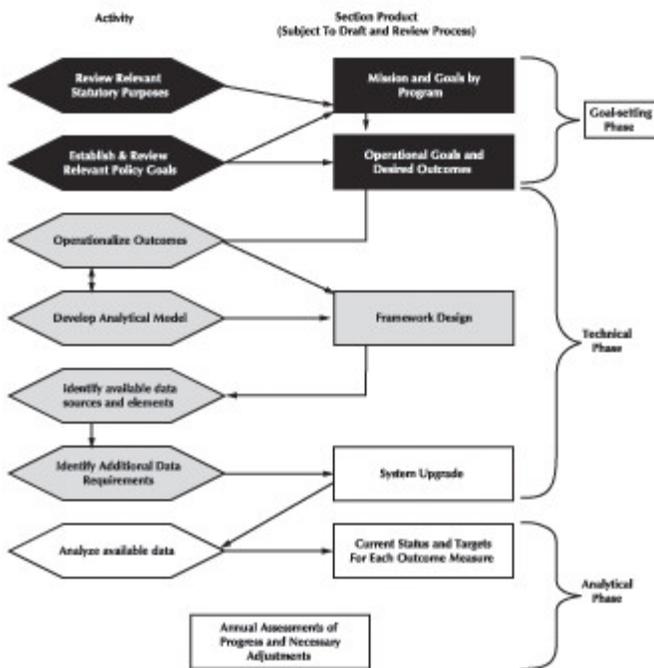
The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

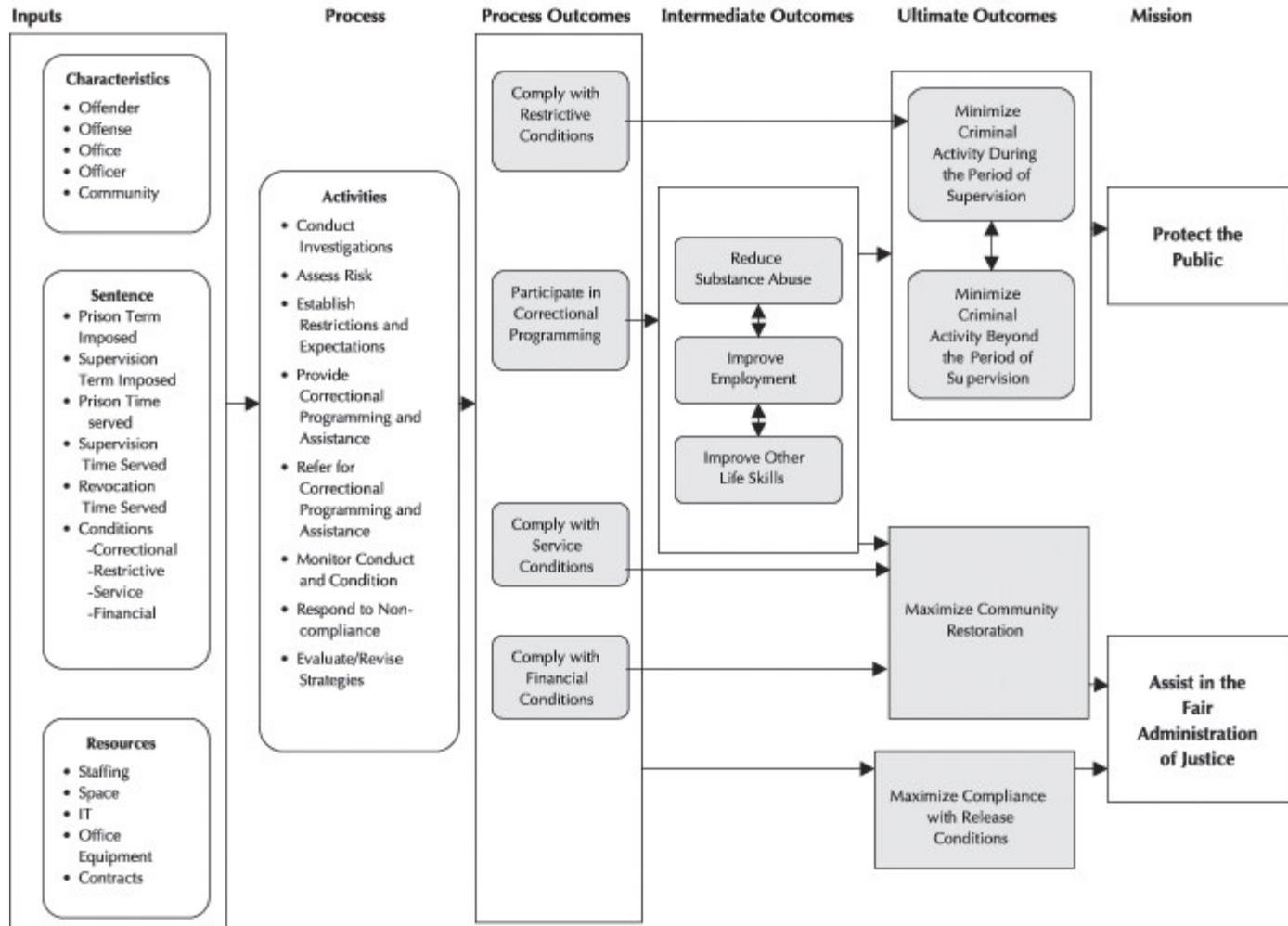
STATUS OF PROCESS FOR ESTABLISHING OUTCOME INDICATORS FOR SUPERVISION FUNCTIONS

FOR SUPERVISION FUNCTIONS

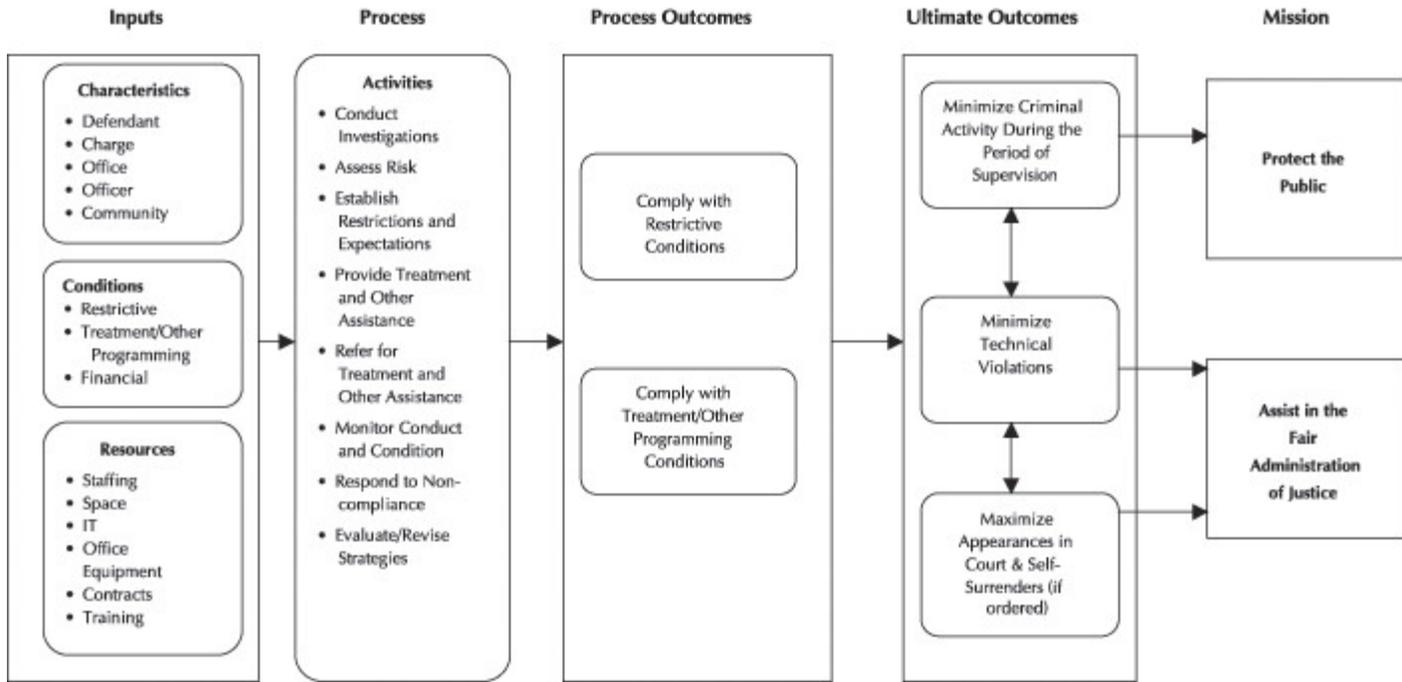
(Black=Completed; Gray=In Progress; White=Future Task)



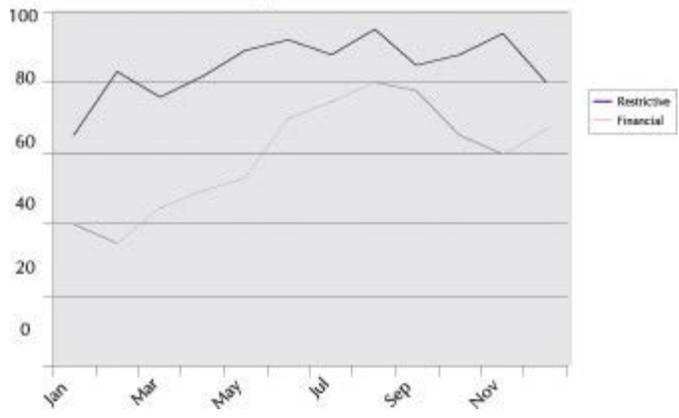
Post-Conviction Supervision



Attachment 4: Pretrial Services Supervision Logic Model



Sample Trend Lines



Contributors To This Issue

James Austin

President, The JFI Institute. Ph.D., University of California, Davis. Co-author of "The Diminishing Returns of Increased Incarceration," *NIC* (2004).

Kimberly Bender

Doctoral student, University of Texas at Austin. M.S.W., Colorado State University. Previously, Outpatient Child and Family Therapist, Fairfax County Community Mental Health Center, Fairfax, VA. Co-author of "Treatment effectiveness with dually diagnosed adolescents: A systematic review," *Brief Treatment and Crisis Intervention*, 6(3) (2006).

James Byrne

Professor, Department of Criminal Justice and Criminology, University of Massachusetts, Lowell . Co-author of *Emerging Roles and Responsibility in the Reentry Partnership Initiative: New Ways of Doing Business* (NIJ 2003).

Terry Fain

Senior Research Associate, RAND Corporation. M.A., Vanderbilt University. Co-author of "Accomplishments in Juvenile Probation in California," *Federal Probation* (June 2006).

Anthony W. Flores

Assistant professor, California State University, Bakersfield . Doctoral candidate, University of Cincinnati . M.S. in Criminal Justice, University of Cincinnati. Has provided technical assistance with offender risk assessment, case planning, and effective rehabilitative programming to correctional agencies in more than 15 states. Author of "Evidence of Professionalism or Quackery: Measuring Practitioner Awareness of Risk/Need Factors and Effective Treatment Strategies," *Federal Probation* (Dec. 2005).

Steve Gottfredson

Provost and Vice President for Academic Affairs, Virginia Commonwealth University. Ph.D., The Johns Hopkins University. Coauthor of "Statistical Risk Assessment: Old Problems and New Applications," *Crime and Delinquency* (January 2006).

Andrew J. Harris

Massachusetts-based public policy consultant in cross-disciplinary approaches to human services and criminal justice issues; visiting faculty member (2006-2007 academic year) in the Criminal Justice Department at the University of Massachusetts-Lowell. Ph.D.,

Patricia Harris

Associate Dean and Professor, College of Public Policy, University of Texas at San Antonio. Ph.D., Rutgers, the State University of New Jersey. Co-author of "Ex-offenders Need Not Apply: The Criminal Background Check in Hiring Decisions," *Journal of Contemporary Criminal Justice* (January 2006).

Edward J. Latessa

Professor and Division Head, University of Cincinnati; consultant with the National Institute of Corrections. Ph.D., Ohio State University. Co-author of *Corrections in America* (2006).

Christopher T. Lowenkamp

Assistant research professor, University of Cincinnati, Division of Criminal Justice; director, The Center for Criminal Justice Research. Previously, Probation Officer, Akron, OH . Ph.D. in Criminal Justice, the University of Cincinnati . Author of “Does Correctional Quality Really Matter?” *Criminology and Public Policy* (August 2006).

Arthur Lurigio

Associate Dean for Faculty, College of Arts and Sciences, Loyola University, Chicago. Previously, Assistant Professor of Psychology, Northwestern University. Ph.D., Loyola University, Chicago. Author of “Taking Stock of Community Corrections Programs, *Criminology and Public Policy*, 4, 259-262 (2005).

Laura Moriarty

Acting Associate Vice Provost for Academic affairs, Virginia Commonwealth University. Ph.D., Sam Houston State University. Coauthor of “Statistical Risk Assessment: Old Problems and New Applications,” *Crime and Delinquency* (January 2006).

April Pattavina

Associate Professor, The University of Massachusetts, Lowell . Editor of and contributing author to *Information Technology and the Criminal Justice System* (2004).

Albert Roberts

Professor of Criminal Justice, Rutgers University. Previously, Associate Professor and Chair, Planning and Management, Curriculum, Indiana University School of Social Work, Indianapolis. D.S.W., University of Maryland .

Paula Smith

Assistant Professor, University of Cincinnati. Ph.D., University of New Brunswick.

James A. Swartz

Associate Professor, Jane Addams College of Social Work, University of Illinois at Chicago . Ph.D., Northwestern University, Chicago . Author of “Screening for Serious Mental Illness Among Criminal Offenders,” *Research in Social Problems and Public Policy* (in press).

Faye Taxman

Professor, Wilder School of Government and Public Policy, Virginia Commonwealth University. Previously, Director of the Bureau of Government Research, University of Maryland, College Park . Ph.D., Rutgers University, Newark, NJ . Author of *Reentry Partnership Initiative* (NIJ).

Susan Turner

Professor of Criminology, Law and Society, University of California, Irvine. Previously, Senior Behavioral Scientist, RAND Corporation. Ph.D., University of North Carolina, Chapel Hill . Co-author of “Accomplishments in Juvenile Probation in California,” *Federal Probation* (June 2006).

[back to top](#)

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

Federal PROBATION

*a journal of correctional
philosophy and practice*

[Home](#)

PUBLISHED BY

The Administrative Office of the United States Court

Leonidas Ralph Mecham, Director

John M. Hughes, Assistant Director
Office of Probation and Pretrial Services

Federal Probation ISSN 0014-9128 is dedicated to informing its readers about current thought, research, and practice in corrections and criminal justice. The journal welcomes the contributions of persons who work with or study juvenile and adult offenders and invites authors to submit articles describing experience or significant findings regarding the prevention and control of delinquency and crime. A style sheet is available from the editor.

Federal Probation is published three times a year, in June, September (on a special topic), and December. Permission to quote is granted on the condition that appropriate credit is given the author and Federal Probation. For information about reprinting articles, please contact the editor.

Subscriptions to Federal Probation are available from the Superintendent of Documents at an annual rate of \$16.00 (\$22.40 foreign). Please see the subscription order form on the last page of this issue for more information.

EDITORIAL STAFF

Timothy P. Cadigan, Executive Editor
Ellen Wilson Fielding, Editor

Federal Probation
Administrative Office of the U.S. Courts
Washington, DC 20544
telephone: 202-502-1600
fax: 202-502-1677

Postmaster: Please send address changes to the editor at the address above.

ADVISORY COMMITTEE

Special Advisor
Merrill A. Smith

Members

Dan Richard Beto
Correctional Management Institute of Texas
Huntsville, Texas

Loren Buddress

Chief Probation Officer
San Mateo County, California

John W. Byrd
San Antonio, Texas

Honorable James G. Carr
United States District Court
Toledo, Ohio

Alvin W. Cohn
Administration of Justice Services, Inc.
Rockville, Maryland

Ronald P. Corbett, Jr.
Executive Director, Supreme Judicial Court
Boston, Massachusetts

Thomas Henry
Newark, New Jersey

Magdeline Jensen
United States Probation Office
Tucson, Arizona

Jolanta Juskiewicz
Pretrial Services Resource Center
Washington, DC

Honorable David D. Noce
United States District Court
St. Louis, Missouri

Joan Petersilia
University of California, Irvine
Irvine, California

Faye Taxman
Virginia Commonwealth University
Richmond, Virginia

[back to top](#)

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

[Home](#)

References

[Endnotes](#)

[Assessment with a Flair: Offender Accountability in Supervision Plans](#)

[What Community Supervision Officers Need to Know About Actuarial Risk Assessment and Clinical Judgment](#)

[Clinical Versus Actuarial Judgments in Criminal Justice Decisions: Should One replace the Other?](#)

[Overcoming Sisyphus: Effective Prediction of Mental Health Disorders and Recidivism Among Delinquents](#)

[Mental Illness in Correctional Populations: The Use of Standardized Screening Tools for Further Evaluation or Treatment](#)

[Risk Assessment and Sex Offender Community Supervision: A Context-Specific Framework](#)

[Validating the Level of Service Inventory—Revised on a Sample of Federal Probationers](#)

[Validation of the Risk and Resiliency Assessment Tool for Juveniles in the Los Angeles County probation System](#)

[Assessing the Role of Clinical and Actuarial Risk Assessment in an Evidence-Based Community Corrections System: Issues to Consider](#)

Assessment with a Flair: Offender Accountability in Supervision Plans

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders (DSM IV) (4th ed.)*. Washington, DC: American Psychiatric Association.

Andrews, D. A., Bonta, J., & Hoge, R. (1990). Classification for Effective rehabilitation: Rediscovering Psychology. *Criminal Justice Behaviour*, 17, 19-52.

Andrews, D. A., Bonta, J., & Wormith, S. J. (2006). The Recent Past and Near Future of Risk and/or Need Assessment. *Crime & Delinquency*, 52 (1), 7-27.

Andrews, D. A., & Bonta, J.L. (1999). *LSI-R:SV Level of Service Inventory Revised: Screening Version User's Manual*. Toronto : Multi-Health Systems.

Andrews, D. A., Zinger, I., Hoge, R., Bonta, J., Gendreau, P., & Cullen, F. (1990). Does Correctional Treatment Work? A Psychologically Informed Meta Analysis. *Criminology*, 28,

Austin, James (2006). How Much Risk Can We Take: The Misuse of Risk Assessment in Corrections. *Federal Probation*, 69 (2), 58-63.

Belenko, S., & Peugh, J. (2005). Estimating drug treatment needs among state prison inmates. *Drug & Alcohol Dependence*, 77(3), 269-281.

DeMatteo, D. S., Marlowe, D. B., & Festinger, D. S. (2006). Secondary Prevention Services for Clients Who Are Low Risk in Drug Court: A Conceptual Model. *Crime & Delinquency*, 52 (1), 114-134.

Gottfredson, M., & Hirschi, T. (1990) *A General Theory of Crime*. Palo Alto, CA: Stanford University Press.

Gottfredson, D. C., Najaka, S. S., & Wilson, D. B. (2001). School-based prevention of problem behaviors: A meta-analysis. *Journal of Quantitative Criminology*, 17 (3), 247-273.

Gottfredson, S.D., & Moriarty, L.J. (2006) Statistical Risk Assessment: Old Problems and New Applications, *Crime & Delinquency*, 52: 178 - 200.

Harris, P. (2006). What Community Supervision Officers Need to Know About Actuarial Risk Assessment and Clinical Judgment. *Federal Probation*, 69 (2), 8-14.

Hare, R. (1999). *Without Conscience: The Disturbing World of the Psychopaths Among Us* . New York: Guilford Press.

Knight, K., Garner, B. R., Simpson, D. D., Morey, J. T., & Flynn, P. M. (2006). An assessment for criminal thinking. *Crime and Delinquency*, 52 (1), 159-177.

Latessa, E. J., Cullen, F. T., & Gendreau, P. (2002) Beyond Correctional Quackery—Professionalism and the possibility of effective treatment. *Federal Probation*, 66, 43-49.

Lipsey, M. W., & Landenberger, N. A. (2006). Cognitive-behavioral interventions. In B. C. Welsh & D. P. Farrington (Eds.), *Preventing crime: What works for children, offender, victims, and places* . Great Britain: Springer.

Lipsey, M. W., & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D. P. Farrington (Eds.). *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions* (pp. 313-345). Thousand Oaks, CA: Sage.

Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2006). The Risk Principle in Action: What Have We Learned From 13,676 Offenders and 97 Correctional Programs? *Crime & Delinquency*, 52 (1), 77-93.

Taxman, F. S. (2004). The Offender and Reentry: Supporting Active Participation in Reintegration. *Federal Probation*, September, 31-35.

Taxman, F. S., & Thanner, M. (2006). Risk, Need, and Responsivity (RNR): It All Depends. *Crime & Delinquency*, 52 (1), 28-51.

Taxman, F. S., Sheperdson, E., & Byrne, J. M. (2004). *Tools of the Trade: A Guide to Incorporating Science into Practice* . Washington, DC: National Institute of Corrections.

Taylor, B. G., Fitzgerald, N, Hunt, D, Reardon, J. A., & Brownstein, H. H. (2001). *ADAM Preliminary 2000 Findings on Drug Use and Drug Markets* (NCJ 189101). Washington, DC: Office of Justice Programs.

Walters, G. (1990). *The Criminal Lifestyle: Patterns of Serious Criminal Conduct* . Sage Publications.

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

measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior*, 32 (5), 479-510.

Ennis, B.J., & Litwack, T.R. (1974). Psychiatry and the presumption of expertise: Flipping coins in the courtroom. *California Law Review*, 62, 693-752.

Glaze, L.E., & Palla, S. (2005). Probation and Parole in the United States, 2004. Bureau of Justice Statistics Bulletin (NCJ 210676). Washington, DC: U.S. Department of Justice.

Glaser, D. (1955). The efficacy of alternative approaches to parole prediction. *American Sociological Review*, 20, 283-287.

Glaser, D. (1962). Prediction tables as accounting devices for judges and parole boards. *Crime and Delinquency*, 8 (3), 239-258.

Goldberg, L.R. (1970). Man versus model of man: A rationale, plus some evidence for a method of improving on clinical inference. *Psychological Bulletin*, 73, 422-432.

Gottfredson, D.M. (1961). Comparing and combining subjective and objective parole predictors. *Research Newsletter*, 3 (Sept-Dec.). Vacaville: California Medical Facility.

Gottfredson, D.M., & Beverly, R.F. (1962). Development and operational use of prediction methods in correctional work. Proceedings of the Social Statistics Section. Washington, DC: American Statistical Association.

Gottfredson, S.D., & Gottfredson, D.M. (1986). Accuracy of prediction models. In Blumstein, A. et al., (eds.), *Criminal Careers and "Career Criminals,"* Vol. II, Washington, DC: National Academy of Sciences Press.

Gottfredson, S.D., & Moriarty, L.J. (2006). Statistical risk assessment: Old problems and new applications. *Crime & Delinquency*, 52 (1), 178-200.

Gough, H.G. (1962). Clinical versus statistical prediction in psychology. In L. Postman (ed), *Psychology in the Making*, pp. 526-584. New York: Knopf.

Hanson, R. K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence*, 20 (2), 212-217.

Harris, G.T., Rice, M.E., & Cormier, C.A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law and Human Behavior*, 26 (4), 377-394.

Hilton, N. Z., Harris, G.T., Rawson, K., & Beach, G.A. (2005). Communicating violence risk information to forensic decision makers. *Criminal Justice and Behavior*, 32 (1), 97-116.

Hilton, N. Z. & Harris, G.T. (2005). Predicting wife assault: A critical review and implications for policy and practice. *Trauma, Violence and Abuse*, 6 (1), 3-23.

Hilton, N. Z., Harris, G.T., & Rice, M.E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *Counseling Psychologist*, 34 (3), 400-409.

Hogarth, R. M. (1980). Judgement and choice: The psychology of decision. New York: John Wiley.

Holland, T.R., Holt, N., Levi, M., & Beckett, G.E. (1983). Comparison and combination of clinical and statistical predictions of recidivism among adult offenders. *Journal of Applied Psychology*, 68 (2), 203-211.

Kleinmuntz, B., Faust, D., Meehl, P.E., & Dawes, R. M. (1990). Clinical versus actuarial judgment. *Science, New Series*, 247 (4939), 146-147.

Litwack, T.R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology Public Policy and Law*, 7 (2), 409-443.

Meehl, P.E. (1965). Seer over sign: The first good example. *Journal of Experimental Research in Personality*, 1, 27-32.

Meehl, P.E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52 (3), 194-216.

Mills, J. F. (2005). Advances in the assessment and prediction of interpersonal violence. *Journal of Interpersonal Violence*, 20 (2), 236-241.

Monahan, J. (1981). *Predicting violent behavior: An assessment of clinical techniques*. Beverly Hills, CA: Sage.

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

Somerville, R.A., & Taffler, R.J. (1995). Banker judgment versus formal forecasting models: The case of country risk assessment. *Journal of Banking & Finance*, 19, 281-297.

Westen, D., & Weinberger, J. (2005). In praise of clinical judgment: Meehl's forgotten legacy. *Journal of Clinical Psychology*, 61 (10), 1257-1276.

[back to top](#)

Overcoming Sisyphus: Effective Prediction of Mental Health Disorders and Recidivism Among Delinquents

Archer, R. P. (1997). *MMPI-A: Assessing adolescent psychopathology* (2 nd Ed.). Mahwah, NJ: Erlbaum.

Archer, R. P., & Baker, E. M. (2005). Minnesota Multiphasic Personality Inventory—Adolescent. In T. Grisso, G. Vincent, & D. Seagrave (Eds.) *Mental health screening and assessment in juvenile justice* (pp. 240-252). New York : Guilford Press.

Archer, R. P., Bolinsky, P. K., Morton, T. L., & Farris, K. L. (2003). MMPI-A characteristics of male adolescents in juvenile justice and clinical treatment settings. *Assessment*, 10, 400-410.

Aronen, E. T., Noam, G. G., & Weinstein, S. R. (1993). Structured diagnostic interviews and clinicians' discharge diagnoses in hospitalized adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 674-681.

Bartel, P., Borum, R., & Forth, A. (2000). *Structured Assessment for Violence Risk in Youth (SAVRY): Consultation Edition*. Tampa : Louis de la Parte Florida Mental Health Institute, University of South Florida.

Borum, R., Bartel, P., Forth, A. (2005). Structural Assessment of Violence Risk in Youth. In T. Grisso, G. Vincent, & D. Seagrave (Eds.) *Mental health screening and assessment in juvenile justice* (pp. 311-323). New York : Guilford Press.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., et al. (1992). *Minnesota Multiphasic Personality Inventory—Adolescent (MMPIA): Manual for administration, scoring and interpretation*. Minneapolis : University of Minnesota Press.

Cambell, M. A., Porter, S., & Santor, D. (2004). Psychopathic traits in adolescent offenders: An evaluation of criminal history, clinical, and psychosocial correlates. *Behavioral Sciences and the Law*, 22, 23-47.

Catchpole, R., & Gretton, H. (2003). The predictive validity of risk assessment with violent

young offenders: A 1-year examination of criminal outcome. *Criminal Justice and Behavior*, 30, 688-708.

Cauffman, E. (2004). A statewide screening of mental health symptoms among juvenile offenders in detention. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43 (4), 430-439.

Cauffman, E. & MacIntosh, R. (2006). A Rasch Differential Item Functioning Analysis of the Massachusetts Screening Instrument. *Educational and Psychological Measurement*, 66 (3), 502-521.

Corrado, R. R., Vincent, G. M., Hart, S. D., & Cohen, I. M. (2004). Predictive validity of the Psychopathy Checklist: Youth Version for general and violent recidivism. *Behavioral Sciences and the Law*, 22, 5-22.

Dembo, R. & Anderson, A. (2005). Problem-Oriented Screening Instrument for Teenagers. In T. Grisso, G. Vincent, & D. Seagrave (Eds.) *Mental health screening and assessment in juvenile justice* (pp. 112-121) . New York : Guilford Press.

Dembo, R., Schmeidler, J., & Walters, W. (2004). Juvenile assessment centers: An innovative approach to identify and respond to youths with substance abuse and related problems entering the justice system. In A. R. Roberts (Ed.). *Juvenile Justice Sourcebook: Past, Present and Future* (pp. 512-536).

Dembo, R., Turner, G., Schmeidler, J., Sue, C. C., Borden, P., & Manning, D. (1996). Development and evaluation of a classification of high risk youths entering a juvenile assessment center. *International Journal of Addictions*, 31, 303-322.

Devine, P., Coolbaugh, K., & Jenkins, S. (1998). Disproportionate minority confinement: Lessons learned from five states. *OJJDP Juvenile Justice Bulletin*, December, 1-11.

Edens, J., Buffington, J., Colwell, K., Johnson, D., Johnson, J. (2002). Psychopathy and institutional misbehavior among incarcerated sex offenders: A comparison of the Psychopathy Checklist-Revised and the Personality Assessment Inventory. *International Journal of Forensic Mental Health*, 1 (1), 49-58.

Espelage, D. L., Cauffman, E., Broidy, L., Piquero, A. R., Mazerolle, P., & Steiner, H. A clusteranalytic investigation of MMPI profiles of serious male and female juvenile offenders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42 (7), 770-777.

Farrington, D. P. (1985). Predicting self-reported and official delinquency. In D. P. Farrington and R. Tarling (eds.). *Prediction in Criminology* (pp. 150-173). Albany, N.Y.: SUNY Press.

Farrington, D. P. & Loeber, R. (2000). Epidemiology of juvenile violence. *Child and Adolescent Psychiatric Clinics of North America* . Vol. 9:733-748.

Farrington, D. P. & West, D. (1990). The Cambridge Study in Delinquent Development: A Long-term Follow-up of 411 London Males. In H.J. Kerner and G. Kaiser, (eds.). *Criminality: Personality, Behavior, and Life History* (Berlin: Springer-Verlag)

Farrington, D. P. & D. J. West (1977). *The Cambridge Study of Delinquent Development* . Cambridge, England: Institute of Criminology, University of Cambridge.

Forbey, J. D. (2003, June). *A review of the MMPIA research literature*. Paper presented at the 38th Annual Symposium on Recent Developments in the Use of the MMPI-2 and MMPI-A, Minneapolis, MN.

Forth, A. E., Kosson, D. S., & Hare, R. D. (2003). *The Hare Psychopathy Checklist: Youth Version* . Toronto: Multi-Health Systems.

- Gretton, H. & Abramowitz, C. (2002, March). *SAVRY: Contribution of items and scales to clinical risk judgments and criminal outcomes*. Paper presented at the biennial conference of the American Psychological-Law Society, Austin, TX.
- Grisso, T. (2005). Evaluating the properties of instruments for screening and assessment. In T. Grisso, G. Vincent, & D. Seagrave (Eds.) *Mental health screening and assessment in juvenile justice* (pp. 71-96) . New York : Guilford Press.
- Grisso, T. & Quinlan, J. C. (2005). Massachusetts Youth Screening Instrument—Version 2. In T. Grisso, G. Vincent, & D. Seagrave (Eds.) *Mental health screening and assessment in juvenile justice* (pp. 71-96) . New York : Guilford Press.
- Grisso, T. & Barnum, R. (2003). *Massachusetts Youth Screening Instrument—Version 2: User's manual and technical report*. Sarasota, FL : Professional Resource Press.
- Grisso, T., Barnum, R., Fletcher, K. E., Cauffman, E., & Peuschold, D. (2001). Massachusetts Youth Screening Instrument for mental health needs of juvenile justice youths. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40 (5), 541-548.
- Hall, J. A., Richardson, B., Spears, J., & Rembert, J. K. (1998). Validation of the POSIT: Comparing drug using and abstaining youth. *Journal of Child and Adolescent Substance Abuse*, 8, 29-61.
- Hathaway, S. R., & Monachesi, E. D. (1963). *Adolescent personality and behavior: MMPI patterns of normal, delinquent, dropout, and other outcomes*. Minneapolis : University of Minnesota Press.
- Hodges, K. (2000a). *Child and Adolescent Functional Assessment Scale* (3rd ed.). Ypsilanti: Eastern Michigan University. Hodges, K. & Wong, M. M. (1997). Psychometric characteristics of a multidimensional measure to assess impairment: The Child and Adolescent Functional Assessment Scale (CAFAS). *Journal of Child and Family Studies*, 5, 445- 467.
- Hoge, R. D., Andrews, D. A., & Leschied, A. W. (2002). *Youth Level of Service/Case Management Inventory: User's manual*. North Tonawanda, N: Multi-Health Systems.
- Knight, J. R., Goodman, E., Pulerwitz, T., & DuRant, R. H. (2001). Reliability of the POSIT in adolescent medial practice. *Journal of Adolescent Health*, 29, 125-130.
- McEachran, A. (2001). *The predictive validity of the PCL-YV and the SAVRY in a population of adolescent offenders*. Unpublished master's thesis, Simon Fraser University, Burnaby, BC, Canada.
- Millon, T. (1993). *Millon Adolescent Clinical Inventory*. Minneapolis, MN : National Computer systems.
- O'Neill, M. L., Lidz, V., & Heilbrun, K. (2003a). Adolescents with psychopathic characteristics in a substance abusing cohort: Treatment process and outcomes. *Law and Human Behavior*, 27, 299-313.
- Piersma, H. L., Pantle, M. L., Smith, A., Boes, J., & Kubiak, J. (1993). The MAPI as a treatment outcome measure for adolescent inpatients. *Journal of Clinical Psychology*, 49, 709-714.
- Rahdert, E. R. (1991). *The Adolescent Assessment/Referral System*. Rockville, MD : National Institute on Drug Abuse.
- Rapp-Palicch, L. & Roberts, A. R. (2004). Mental illness and juvenile offending. In A. R. Roberts (editor). *Juvenile Justice Sourcebook: Past, Present and Future*. Pp. 289-308.
- Roberts, A. R. (2004). The emergence of the juvenile court and probation services. In A. R. Roberts (editor). *Juvenile Justice Sourcebook: Past, Present and Future*. Pp. 163-181.

- Rowe, R. (2002). *Predictors of criminal offending: Evaluating measures of risk/needs, psychopathy, and destructive behavior disorders*. Unpublished doctoral dissertation, Carleton University, Ottawa, ON, Canada.
- Salekin, R. T., Leistico, A. R., Schrum, C. L., & Mullins, J. (2005). Millon Adolescent Clinical Inventory. In T. Grisso, G. Vincent, & D. Seagrave (Eds.) *Mental health screening and assessment in juvenile justice* (pp. 253-264). New York : Guilford Press.
- Salekin, R. T., Ziegler, T. A., Larrea, M. A., Anthony, V. L., & Bennet, A. D. (2003). Predicting dangerousness with two Millon Adolescent Clinical Inventor Psychopathy Scales: The importance of egocentric and callous traits. *Journal of Personality Assessment, 80*, 154-163.
- Schmid, F., Hoge, R. D., & Robertson, L. (2002, May). *Assessing risk and need in youthful offenders*. Paper presented at the annual conference of the Canadian Psychological Association, Vancouver, BC, Canada.
- Shaffer, D. M., Fisher, P. W., Lucas, C., Dulcan, M. K., & Swab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy Child and Adolescent Psychiatry, 39*, 28-38.
- Sheras, P. (2000). Depression and suicide in juvenile offenders. In *Juvenile Justice Fact Sheet*. Charlottesville: Institute of Law, Psychiatry, and Public Policy, University of Virginia.
- Spain, S. E., Douglas, K. S., Poythress, N. G., & Epstein, M. (2004). The relationship between psychopathic features, violence, and treatment outcome: The comparison of three youth measures of psychopathic features. *Behavioral Sciences and the Law, 22*, 85-102.
- Stewart, D., & Trupin, E. (2003). Clinical utility and policy implications of a statewide mental health screening process for juvenile offenders. *Psychiatric Services, 54* (3), 377-382.
- Taylor, J., Skubic-Kemper, T., Loney, B. R., and Kistner, J. A. (2006). Classification of severe male juvenile offenders using the MACI Clinical and Personality Scales. *Journal of Clinical and Adolescent Psychology, 35* (1), 90-102.
- Teplin, L. A., Abram, K. M., McClelland, G. M., Washburn, J. J., & Pikus, A. K. (2005). Detecting mental disorder in juvenile detainees: Who receives services. *American Journal of Public Health, 95* (10), 1773-1780.
- Teplin, L. A., Abram, K. M., McClelland, G. M., Duclan, M., & Mericle, A. A. (2003). Psychiatric disorders in youth in juvenile detention. *Archive of General Psychiatry, 59*, 1133-1143.
- Wasserman, G. A., McReynolds, L. S., Fisher, P., & Lucas, C. P. (2005). Diagnostic Interview Schedule for Children: Present State Voice Version. In T. Grisso, G. Vincent, & D. Seagrave (Eds.) *Mental health screening and assessment in juvenile justice* (pp. 224-239). New York : Guilford Press.
- Wasserman, G. A. McReynolds, L. S., Ko, S. J., Katz, L. M., Cauffman, E., Haxton, W., et al., (2004). Screening for emergent risk and service needs among incarcerated youth: Comparing MAYSI-2 and Voice DISC-IV. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*, 752-761.
- Wasserman G. A., McReynolds L. S., Ko, S. J., Katz L. M., & Schwank, J. (2005). Gender differences in psychiatric disorder for youths in juvenile probations. *American Journal of Public Health, 95* (1), 131-137.
- Wasserman, G. A., Jensen, P. S., Ko, S. J., Coccozza, J., Trupin, E., Angold, A., Cauffman, E., & Grisso, T. (2003). Mental health assessments in juvenile justice: Report on the Consensus

Conference. *Journal of American Academy of Child and Adolescent Psychiatry*, 42 (7), 752-761.

Wolfgang, M. E., Figlio, R. M. & Sellin, T. (1972). *Delinquency in a birth cohort* (Chicago: the University of Chicago Press).

[back to top](#)

Mental Illness in Correctional Populations: The Use of Standardized Screening Tools for Further Evaluation or Treatment

Abram, K. M., & Teplin, L. A. (1991). Cooccurring disorders among mentally ill jail detainees: Implications for public policy. *American Psychologist*, 46, 1036-1045.

Abram, K. M., Teplin, L. A., & McClelland, G. M. (2003). Comorbidity of severe psychiatric disorders and substance use disorders among women in jail. *American Journal of Psychiatry*, 160, 1007-1010.

Abramson, M.F. (1972). The criminalization of mentally disordered behavior: Possible sideeffect of a new mental health law. *Hospital and Community Psychiatry*, 23, 101-107.

American Psychiatric Association [APA] (1989). *Psychiatric services in jails and prisons. Report of the task force on psychiatric services in jails and prisons*. Washington, DC :

American Psychiatric Association. American Psychiatric Association [APA]. (2002). *Diagnostic and Statistical Manual of Mental Disorders, Text-Revision* . Washington, DC: American Psychiatric Association.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* . Upper Saddle River, NJ: Prentice Hall.

Boone, H. (1995). Mental illness in probation and parole populations: Results from a national survey. *Perspectives*, 28, 32-39.

Diamond, P.M., Wang, E. W., Holzer III, C. E., Thomas, C., & des Anges, C. (2001). The prevalence of mental illness in prison. *Administration and Policy in Mental Health*, 29, 21-40.

Ditton, P. (1999). *Mental health and treatment for inmates and probationers*. Washington, DC: Bureau of Justice Statistics.

Draine, J. (2003). Where is the illness in the criminalization of the mentally ill? In W.H. Fisher (Ed.). *Community-based interventions for criminal offenders with severe mental illness* . New York: Elsevier.

Earley, P. (2006). *Crazy: A father's search through America 's mental health madness* . New York: G.P. Putnam's Sons.

Elliot, R. L. (1997). Evaluating the quality of correctional mental health services: an approach to surveying a correctional mental health system. *Behavioral Science and the Law*, 15, 427-38.

Fisher W, Silver, E. & Wolff, N. (in press). Beyond "Criminalization": Developing a criminologically-informed framework for mental health policy and services research. *Administration and Policy in Mental/Mental Health Services Research* .

Hepburn, J. R. (1994). Classifying drug offenders for treatment. In D. L. MacKenzie & C. D. Uchida (Eds.), *Drugs and crime: Evaluating public policy initiatives* (pp. 172-87). Thousand Oaks, CA : Sage.

Hornung, C. A., Anno, B. J., Greifinger, R. B., & Gadre, S. (2002). Health care for soon-to-be-released inmates: A survey of state prison systems. In *The health status of soon-to-be-released inmates: A report to Congress* (Volume 2, pp. 1-11). Chicago: National Commission on Correctional Health Care.

Human Rights Watch [HRW] (2003). *Illequipped: U.S. prisons and offenders with mental illnesses*. Retrieved October 22, 2003 from <http://www.hrw.org/reports/2003/usa1003/index.htm>.

Insel, T. R. (2003, March). *Introductory presentation*. Presented at Beyond the clinic walls: Expanding mental health, drug and alcohol services research outside the specialty care system, a conference co-sponsored by the National Institute of Mental Health and the National Institute on Alcoholism and Alcohol Abuse, Washington, DC.

Jordan, B. K., Federman, E. B., Burns, B. J., Schlenger, W. E., Fairbank, J. A., & Caddell, J. M. (2002). Lifetime use of mental health and substance abuse treatment services by incarcerated women felons. *Psychiatric Services, 53*, 317-325.

Kessler, R. C., Andrews, G., Mroczek, D., Ustun, T. B., Wittchen, H. U. (1998). The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI-SF). *International Journal of Methods in Psychiatric Research, 7*, 171-185.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. -L., T. et al. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32*, 959-976.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hirpi, E. et al. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry, 60*, 184-189.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H. U. et al. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the national comorbidity study. *Archives of General Psychiatry, 51*, 8-19.

Lamb, H. R., & Weinberger, L. E. (1998). Persons with severe mental illness in jails and prisons: A review. *Psychiatric Services, 49*, 483-492.

Lamb, H. R., Weinberger, L. E., & Gross, B. H. (2004). Mentally ill in the criminal justice system: Some perspectives. *Psychiatric Quarterly, 75*, 107-126.

Lurigio, A. J. (2003). Benefits of integrated treatment for co-occurring substance use disorders and mental illness. *GLATTC Research Update, 5*, 1-2.

Lurigio, A. J. (2005). From intensive care to intensive supervision: The historical factors associated with the growing numbers of mentally ill persons in the criminal justice system. *Public Service Psychology, 30*, 22-23.

Lurigio, A. J., Cho, Y. I., Swartz, J. A., Johnson, T. P., Graf, I., & Pickup, L. (2003). Standardized assessment of substance-related, other psychiatric, and comorbid disorders among probationers. *International Journal of Offender Therapy and Comparative Criminology, 47*, 630-653.

Lurigio, A. J., & Swartz, J. A. (2000). Changing the contours of the criminal justice system to meet the needs of persons with serious mental illness. In J. Horney (Ed.), *NIJ 2000 Series: Policies, processes, and decisions of the criminal justice system* (Vol. 3) (pp. 45-108). Washington, DC: National Institute of Justice.

Lurigio, A.J. & Swartz, J.A. (2006). *Probation officers' perceptions of mental illness among persons on their caseloads*. Chicago: Cook County Adult Probation Department.

Murphy, J. M. (2002). Symptom scales and diagnostic schedules in adult psychiatry. In M. T. Tsuang & M. Tohen (Eds.), *Textbook in psychiatric epidemiology* (2 nd ed. pp. 273- 332). New York: John Wiley & Sons.

National Commission on Correctional Health Care [NCCHC] (2002). *The health status of soon-to-be-released inmates: A report to Congress* (Volumes 1 and 2). Chicago: Author.

Nolen-Hoeksema, S. (2005). *Abnormal psychology*. NY: McGraw-Hill. Peters, R. H., Greenbaum, P. E., Steinberg, M. L., Carter, C. R., Ortiz, M. M., Fry, B. C., &

Valle, S. K. (2000). Effectiveness of screening instruments in detecting substance use disorders among prisoners. *Journal of Substance Abuse Treatment, 18*, 349-58.

Regier, D. A., Kaelber, C. T., Rae, D. S., Farmer, M. E., Knauper, B., Kessler, R. C., & Norquist, G. S. (1998). Limitations of diagnostic criteria and assessment instruments for mental disorders. *Archives of General Psychiatry, 55*, 109-115.

Regier, D. A., Farmer, M. E., Rae, D. A., Locke B. Z., Keith, S. J., Judd, L. L., & Goodwin, F. K. (1990). Comorbidity of mental disorders with alcohol and other drug abuse: Results from the Epidemiological Catchment Area (ECA) Study. *Journal of the American Medical Association, 264*, 2511-2518.

Robins, L. N., Wing, J., Wittchen, H. U., Helzer, J. E., Babor, T. F., Burke, J., et al. (1988). The Composite International Diagnostic Interview. *Archives of General Psychiatry, 45*, 1069-1077.

Sacks, S., Melnick, G., & Coen, C. (2005). *Cooccurring disorders screening instrument for criminal justice populations* (CJ-CODSI). Washington, DC : National Institute on Drug Abuse.

Schmitz, N., Kruse, J., & Tress, W. (2000). Application of stratum-specific likelihood ratios in mental health screening. *Social Psychiatry and Psychiatric Epidemiology, 35*, 375-379.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The Mini-International Neuropsychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview for the DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*, 22-33.

Skeem, J. L., Encandela, J., & Louden, J. E. (2003). Perspectives on probation and mandated mental health treatment in specialized and traditional probation departments. *Behavioral Science and the Law, 21*, 429-58.

Slade, T. B., & Andrews, G. (2002). Empirical impact of DSM-IV diagnostic criterion or clinical significance. *The Journal of Nervous and Mental Disease, 190*, 334-337.

Solomon, P., & Draine, J. (1999). Using clinical and criminal involvement factors to explain homelessness among clients of a psychiatric probation and parole service. *Psychiatric Quarterly, 70*, 75-87.

Steadman, H.J., Scott, J. E., Osher, F., Agnese, T.K., & Robins, P.C. (2005). Validation of the brief jail mental health screen. *Psychiatric Services, 56*, 816-822.

Steadman, H.J., & Veysey, B. M. (1997). *Providing services for jail inmates with mental disorders* (Research in Brief, U.S. DOJ Publication No. NCJ 162207). Washington, DC: U.S. Department of Justice, National Institute of Justice. September 2006

Swartz, J. A. (2001). Considering psychiatric comorbidities among addicted offenders: A new strategy for client-treatment matching. *Offender Substance Abuse Report, 1*, 65-66, 70-73.

Swartz, J. A., & Lurigio, A. J. (2005). Screening for serious mental illness among criminal offenders. *Research in Social Problems and Public Policy, 12*, 137-161.

Swartz, J. A., & Lurigio, A. J. (in press). Serious mental illness and arrest: The generalized mediating effects of substance use. *Crime and Delinquency*.

Teplin, L. A. (1990). Detecting disorder: the treatment of mental illness among jail detainees.

Journal of Consulting and Clinical Psychology, 58, 233-6.

Teplin, L. A., Abram, K. M., & McLelland, G. M. (1997). Mentally disordered women in jail: who receives services? *American Journal of Public Health*, 87, 604-9.

Teplin, L., A., & Swartz, J. A. (1989) . Screening for severe mental disorder in jails: The development of the Referral Decision Scale . *Law and Human Behavior*, 13, 1 - 18.

Torrey, E.F. (1995). Jails and prisons: America's new mental hospitals. *American Journal of Public Health*, 85, 1611-1613.

Veysey, B. M., & Bichler-Robertson, G. (2002, April). Providing psychiatric services in correctional settings. In *The health status of soon-to-be-released inmates: A report to Congress* (Volume 2, pp. 157-165.). Chicago: National Commission on Correctional Health Care.

[back to top](#)

Risk Assessment and Sex Offender Community Supervision: A Context-Specific Framework

Adams, D. (2002). *Summary of State Sex Offender Registries, 2001* . Washington DC: Bureau of Justice Statistics.

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 (4), 490-521.

Boer, D., Hart, S., Kropp, P., & Webster, C. (1997). *Manual for the sexual violence risk-20: Professional guidelines for assessing risk of sexual violence* . Burnaby, British Columbia: Mental Health Law & Policy Institute, Simon Fraser University.

Craissati, J., & Beech, A. (2003). A review of dynamic variables and their relationship to risk prediction in sex offenders. *Journal of Sexual Aggression*, 9 (1), 41-55.

Doren, D. M. (2004a). *Bibliography of Published Works Relative to Risk Assessment for Sexual Offenders* . Retrieved October 15, 2005, from the World Wide Web:
<http://www.atsa.com/pdfs/riskAssessmentBiblio.pdf>.

Doren, D. M. (2004b). Toward a Multidimensional Model for Sexual Recidivism Risk. *Journal of Interpersonal Violence*, 19 (8), 20p.

Dvoskin, J., & Heilbrun, K. (2001). Risk Assessment and Release Decision-Making: Toward Resolving the Great Debate. *J Am Acad Psychiatry Law*, 29, 6-10.

English, K., Pullen, S., & Jones, L. (1997). *Managing Adult Sex Offenders in the Community: A Containment Approach* (163387). Washington, DC: National Institute of Justice.

Epperson, D., Kaul, J., Huot, S., Hesselton, D., Alexander, W., & Goldman, R. (1999). *Minnesota Sex Offender Screening Tool -Revised* . St. Paul: Minnesota Department of Corrections.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34, 401-433.

Greenfeld, L. A. (1997). *Sex offenses and offenders* . Washington DC: Bureau of Justice Statistics Monograph #NCJ-163392, U.S. Department of Justice.

Hanson, R. K. (1997). *The development of a brief actuarial scale for sex offender recidivism (User Report No. 1997-04)*. Ottawa, Canada: Department of the Solicitor General of Canada.

Hanson, R. K., & Bussiere, M. T. (1998). Predicting relapse: a meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66, 3 48-362.

Hanson, R. K., & Harris, A. J. R. (1998). *Dynamic predictors of sexual recidivism* . Ottawa, Canada: Department of the Solicitor General of Canada.

Hanson, R. K., & Harris, A. J. R. (2000a). *The Sex offender needs assessment rating (SONAR): A method for measuring change in risk levels* . Department of the Solicitor General of Canada. Retrieved June 7, 2006, 2006, from the World Wide Web: http://ww2.psepc-sppcc.gc.ca/publications/corrections/pdf/200001b_e.pdf.

Hanson, R. K., & Harris, A. J. R. (2000b). Where should we intervene? Dynamic predictors of sex offense recidivism. *Criminal Justice and Behavior*, 27, 6-35.

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

Hanson, R. K., & Morton-Bourgon, K. (2004). *Predictors of Sexual Recidivism: An Updated Meta-Analysis* . Ottawa, Canada: Public Safety and Emergency Preparedness Canada.

Hanson, R. K., Stey, R. A., & Gauthier, R. (1993). Long-term recidivism of child molesters. *Journal of Consulting and Clinical Psychology*, 61, 646-652.

Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders (User Report No. 99-02)* . Ottawa, Canada: Department of the Solicitor General of Canada.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24 (1), 119-136.

Harris, A. (2005). *Civil Commitment of Sexual Predators: A Study in Policy Implementation* . New York: LFB Scholarly Publishing.

Harris, A. J. R., & Hanson, R. K. (2003, August 2003). The Dynamic Supervision Project: Improving the Community Supervision of Sex Offenders. *Corrections Today* .

Hart, S. D. (1998). The role of psychopathy in assessing risk for violence: conceptual and methodological issues. *Legal Criminal Psychol*, 3, 123-140.

Hart, S. D., Kropp, P. R., & Laws, R. L. (2004). *The Risk for Sexual Violence Protocol (RSVP)* . Burnaby, BC: Mental Health, Law, and Policy Institute, Simon Fraser University.

Hart, S. D., Webster, C., & Menzies, R. (1993). Note on Portraying the Accuracy of Violence Predictions. *Law and Human Behavior*, 17, 695-696.

Hart, T. & Rennison, C. (2003). *Reporting Crime to the Police* . Washington DC: U.S. Department of Justice, Bureau of Justice Statistics.

Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior*, 21, 347-359.

Hepburn, J., & Griffin, M. (2004). *An Analysis of Risk Factors Contributing to the Recidivism of Sex Offenders on Probation* . Washington DC: National Institute of Justice.

Hughes, T., Wilson, D., & Beck, A. (2001). *Trends in State Parole, 1990-2000* . Washington DC: Bureau of Justice Statistics.

Huot, S. (1999). The Referral Process, *Schlank, Anita (Ed); Cohen, Fred (Ed). (1999). The sexual predator: Law, policy, evaluation and treatment.* (pp. 6-1-6-10). Kingston, NJ, US: Civic Research Institute.

Kern, R. P., & Farrar-Owens, M. (2004). Sentencing Guidelines with Integrated Offender Risk Assessment. *Federal Sentencing Reporter*, 16 (3), 165-169.

Knight, R. A. (1999). Validation of a typology for rapists. *Journal of Interpersonal Violence*, 14, 303-330.

Knight, R. A., Carter, D., & Prentky, R. A. (1989). A system for the classification of child molesters: Reliability and application. *Journal of Interpersonal Violence*, 4, 3-23.

Knight, R. A., & Prentky, R. A. (1990). Classifying Sexual Offenders: The Development and corroboration of taxonomic models. In W. L. Marshall & D. R. Laws & H. E. Barbaree (Eds.), *Handbook of Sexual Assault: Issues, Theories, and Treatment of the Offender*. New York: Plenum.

Lanning, K. (1986). *Child Molesters: A Behavioral Analysis for Law Enforcement Officers Investigating Cases of Sexual Exploitation*. Washington DC: National Center for Missing and Exploited Children.

Meloy, M. (2005). The Sex Offender Next Door: An Analysis of Recidivism, Risk Factors, and Deterrence of Sex Offenders on Probation. *Criminal Justice Policy Review*, 16 (2), 211-236.

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.

National Conference of State Legislatures. (2006). *State Crime Legislation in 2005*. National Conference of State Legislatures. Retrieved June 18, 2006, 2006, from the World Wide Web: <http://www.ncsl.org/programs/cj/05crime.htm>.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. (1998). *Violent Offenders: Appraising and Managing Risk*. Washington, D.C: American Psychological Association.

Schlesinger, L. (2004). *Sexual Murder: Catathymic and Compulsive Homicides*. Boca Raton, Florida: CRC Press.

Seto, M. C. (2005). Is More Better? Combining Actuarial Risk Scales to Predict Recidivism Among Adult Sex Offenders. *Psychological Assessment*, 17 (2), 156-167.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1, 1-26.

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

[back to top](#)

Validating the Level of Service Inventory—Revised on a Sample of Federal Probationers

Andrews, D., & Bonta, J. (1995). *LSI-R: The Level of Service Inventory-Revised*. Toronto, ONT: Multi-Health Systems, Inc.

Andrews, D., & Bonta, J. (2003). *The psychology of criminal conduct*. Cincinnati: Anderson.

Andrews, D. A., Kiessling, Robinson, D., & Mickus, S. G. (1986). The risk principle of case classification: An outcome evaluation with young adult probationers. *Canadian Journal of Criminology*, 28 (4), 377-384.

Andrews, D., Zinger, I., Hoge, R., Bonta, J., Gendreau, P., & Cullen, F. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis.

Criminology, 28 (3), 369-404.

Baird, C., Hines, C., and Bemus, B. (1979). *The Wisconsin Case Classification/Staff Deployment Project: Two-Year Follow-Up Report*. Madison, WI: Wisconsin Division of Corrections.

Bonta, J. L. (1989). Native inmates: Institutional response, risk, and needs. *Canadian Journal of Criminology*, 39, 49-62.

Bonta, J., & Motiuk, L. L. (1985). Utilization of an interview-based classification instrument: A study of correctional halfway houses. *Criminal Justice and Behavior*, 12 (3), 333-352.

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

Brown, S. L. (2002). The dynamic prediction of criminal recidivism: A three wave prospective study. Unpublished dissertation, Queen's University, Kingston, Ontario, Canada.

Clements, C. B. (1996). Offender classification: Two decades of progress. *Criminal Justice and Behavior*, 23 (1), 121-143.

Cullen, F. T., Wright, J. P., & Applegate, B. K. (1996). Control in the community: The limits of reform? In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 69-116). Newbury Park, CA: Sage.

Dowden, C., & Andrews, D.A. (1999a). What works for female offenders: A meta-analytic review. *Crime and Delinquency*, 45 (4), 438-452.

Dowden, C., & Andrews, D.A. (1999b). What works in young offender treatment: A metaanalysis. *Forum on Corrections Research*, 11, 21-24.

Dowdy, E. R., Lacy, M. G., & Unnithan, N. P. (2001). Correctional prediction and the Level of Service Inventory. *Journal of Criminal Justice*, 30, 29-39.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works? *Criminology*, 34, 575-607.

Glaze, L. E., & Palla, S. (2005). *Probation and parole in the United States, 2004* (NCJ 210676). Bureau of Justice Statistics Bulletin. Washington, DC: U.S. Department of Justice.

Goggin, C., Gendreau, P., & Gray, G. (1998). Associates and social interaction. *Forum on Corrections Research*, 10 (3), 24-27.

Gottfredson, S., & Gottfredson, D. (1986). The accuracy of prediction models. In A. Blumstein, J. Cohen, & C. Visher (Eds.), *Criminal careers and "career criminals"* (pp. 212-290). Albany: SUNY.

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 (1), 19-30.

Holsinger, A. M., Lowenkamp, C. T., & Latessa, E. J. (2004). Validating the LSI-R on a sample of jail inmates. *Journal of Offender Monitoring*, Winter/Spring, 8-9.

Jones, D. A., Johnson, S., Latessa, E. J., & Travis, L. F. (1999). Case classification in community corrections: Preliminary findings from a national survey. Topics in Community Corrections, National Institute of Corrections, U.S. Department of Justice, Washington, D.C.

Latessa, E. (2003-2004). Best practices of classification and assessment. *Journal of Community Corrections*. Winter 2003-2004.

Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2001). Risk/need assessment, offender

classification, and the role of child abuse. *Criminal Justice and Behavior*, 28 (5), 543-563.

Lowenkamp, C.T. & Latessa, E. J. (2004a). Increasing the effectiveness of correctional programming through the risk principle: Identifying offenders for residential placement. *Criminology and Public Policy*, 4 (1), 501-528.

Lowenkamp, C.T. & Latessa, E. J. (2004b). Residential community corrections and the risk principle: Lessons learned in Ohio. *Ohio Corrections Research Compendium*, 2, 245-254 .

Lowenkamp, C.T. & Latessa, E. J. (2005). The role of offender assessment tools and how to select them . *For the Record*, 4th quarter, 18-20,

Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis, MN: University of Minnesota Press.

Monahan, J. (1981). *Predicting violent behavior*. Beverly Hills, CA : Sage.

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

Motiuk, L. L., Bonta, J., & Andrews, D.A. (1986). Classification in correctional halfway houses: The relative and incremental predictive criterion validities of the Megargee-MMPI and LSI systems. *Criminal Justice and Behavior*, 13-46.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive accuracy. *Journal of Consulting and Clinical Psychology*, 63 (5), 737-748.

Taylor, G. (2001). Importance of developing correctional plans for offenders. *Forum on Corrections Research*, 13, 14-17.

United States Probation and Pretrial Services (2005). *Year-in-review report: Fiscal year 2004* . Washington D.C.: Administrative Office of the United States Courts.

Wright, K., Clear, T., & Dickson, P. (1984). Universal application of probation risk assessment instruments: A critique. *Criminology*, 22 (1), 113-134.

[back to top](#)

Validation of the Risk and Resiliency Assessment Tool for Juveniles in the Los Angeles County probation System

Bonta, J. (1996). Risk-needs assessment and treatment. In A. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* . Thousand Oaks, CA: Sage.

Jones, D., Johnson, S., Latessa, E., & Travis, L. F. (1999). Case classification in community corrections: Preliminary findings from a national survey in New York state. In LIS, Inc., National Institutes of Corrections, *Topics in community corrections*, pp. 4-10.

Little, J. (n.d.). An evaluation of the San Diego risk and resiliency check up. Boulder, CO: Social Science Data Analysis Center, Institute of Behavioral Science, University of Colorado, Boulder.

Multi-Health Systems Inc. (1998). The Youth Level of Service/Case Management Inventory (YLS/CMI): Research Version. North Tonawanda, NY: Multi-Health Systems Inc.

Solomon, L. & Camp, A. T. (1993). The revolution in correctional classification. In American Correctional Association (Ed.), *Classification: A tool for managing today's offenders*. Laurel, MD: American Correctional Association.

Turner, S., & Fain, T. (2003). Validation of the Los Angeles County Probation Department's risk

and needs assessment instruments. Santa Monica, Calif.: RAND Corporation.

Washington State Institute for Public Policy. (2004). Assessing risk for reoffense: Validating the Washington State Juvenile Court Assessment. Olympia, WA: Washington State Institute for Public Policy.

[back to top](#)

Assessing the Role of Clinical and Actuarial Risk Assessment in an Evidence-Based Community Corrections System: Issues to Consider

Brumbaugh, Susan and Danielle Steffey (2005). "The Importance of Constructing and Validating Risk Assessment Instruments in Community Corrections Settings," *Justice Research and Policy* 7(2):57-86.

Byrne, James and Faye Taxman (2006). "Crime Control Strategies and Community Change," *Federal Probation* June: 3-12.

Byrne, James and Faye Taxman (2005) "Crime (Control) is a Choice: Divergent Perspectives on the Role of Treatment in the Adult Corrections System," *Criminology and Public Policy* 4(2):291-310.

Byrne, James, Robin Robinson, and Anthony Braga (1992). "An Examination of Gender Bias in the Classification and Treatment of Juvenile Offenders," unpublished manuscript.

Byrne, James and Robin Robinson (1990). *Juvenile Risk Classification in Illinois: An Examination of Inter-Rater Reliability* (Springfield, Illinois: Administrative Office of the Illinois Courts).

Gottfredson, Don and Michael Tonry (1987). *Prediction and Classification: Criminal Justice Decision-Making*. Chicago, University of Chicago Press.

Gottfredson, Steven and Robert Taylor (1986). "Person-Environment Interactions in the Prediction of Recidivism," pp 133-155, in James Byrne and Robert Sampson, editors. *The Social Ecology of Crim*. New York: Springer-Verlag.

Jacobson, Jerry O. (2006). "Do Drug Treatment Facilities Increase Clients' Exposure to Potential Neighborhood-Level Triggers for Relapse? A Small Area Assessment of a Large Public Treatment System" *Journal of Urban Health* (March):1-13.

Pattavina, April and Faye Taxman (forthcoming). "Community Corrections and Soft Technology" in James Byrne and Don Rebovich, editors, *The New Technology of Crime, Law, and Social Control* (New York: Criminal Justice Press).

Pattavina, April, James Byrne and Luis Garcia (2006). "An Examination of Citizen Involvement in Crime Prevention in High Risk Versus Low to Moderate Risk Neighborhoods" *Crime and Delinquency* 52(2):203-231.

Sampson, Robert and Lydia Bean (2005). "Cultural Mechanisms and Killing Fields: A Revised Theory of Community-Level Racial Inequality" in Peterson, Krivo, and Hagan, editors, *The Many Colors of Crime: Inequalities of Race, Ethnicity and Crime in America*. New York: New York University Press (retrieved from Robert Sampson's webpage).

Sampson, Robert J, Jeffrey Morenoff, and Stephen Raudenbush (2005). "Social Anatomy of Racial and Ethnic Disparities in Violence," *American Journal of Public Health* 95(2):224-232.

Sampson, Robert J. and Stephen W. Raudenbush (2004). "Seeing Disorder: Neighborhood Stigma and the Social Construction of 'Broken Windows,'" *Social Psychology Quarterly* 67(4):319-342.

Sampson, Robert J. and Stephen W. Raudenbush (2001). "Disorder in Urban Neighborhoods — Does it Lead to Crime?" *Research in Brief* 1-5 (February).

Welsh, Brandon and David Farrington (2006). "Evidence-Based Crime Prevention" pp 1-20 in Welsh and Farrington, Editors *Preventing Crime: What Works for Offenders, Victims, and Places* . New York: Springer.

[back to top](#)

[Endnotes](#)

The articles and reviews that appear in *Federal Probation* express the points of view of the persons who wrote them and not necessarily the points of view of the agencies and organizations with which these persons are affiliated. Moreover, *Federal Probation's* publication of the articles and review is not to be taken as an endorsement of the material by the editors, the Administrative Office of the U.S. Courts, or the Federal Probation and Pretrial Services System.

Published by the Administrative Office of the United States Courts www.uscourts.gov
[Publishing Information](#)

[Home](#)

Endnotes

[References](#)

[Clinical Versus Actuarial Judgments in Criminal Justice Decisions: Should One replace the Other?](#)

[Validation of the Risk and Resiliency Assessment Tool for Juveniles in the Los Angeles County probation System](#)

[How Much Risk Can We Take? The Misuse of Risk Assessment in Corrections](#)

[From Theory to Practice: The Lifecycle Document for the Results-Based Management Framework for the Federal Probation and Pretrial Services System](#)

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

¹ Portions of the preceding discussion adapted from Gottfredson and Gottfredson (1986).

² Historical-Clinical-Risk-Management-20 (HRC-20).

[back to top](#)

Validation of the Risk and Resiliency Assessment Tool for Juveniles in the Los Angeles County probation System

¹ These estimates are based on unpublished analyses from Turner & Fain (2003).

² Police refer cases to the District Attorney in Los Angeles for processing. Youths charged with offenses for which the District Attorney must file and those youths who are detained in juvenile hall are directed to the Court for arraignment. The SDRRC was administered at this pre-plea stage for these “court” cases. Youths not initially referred to court—those generally with more minor offenses—are referred to Probation to make a determination of how to handle the case. These “non-court” cases can received a number of possible outcomes, including having the case closed, the youth being placed on informal probation, or the case being referred to court. The SDRRC was administered to “noncourt” cases at this point.

³ Due to logistical restrictions, we were not able to pilot test the SDRRC in all area Probation Department offices.

⁴ For weighting purposes, ethnicity was divided into five categories: black, Hispanic, white, other race, and unknown. Age was categorized as less than 13, greater than 18, and single years of age for ages 13-18.

⁵ We were not able to use age in calculating weights because there were too few

representatives in the sample for some combinations of gender, race/ethnicity, court type, and age.

⁶ This includes those with negative scores.

⁷ We include the square of age as a factor in the logistic regression because age has a curvilinear relationship with rearrest. Little (n.d.) used a similar analytic approach in her evaluation of the SDRRC.

[back to top](#)

How Much Risk Can We Take? The Misuse of Risk Assessment in Corrections

¹ See <http://bmj.bmjournals.com/cgi/content/full/312/7023/71>. Also see <http://www.ahrq.gov/clinic/epc/> for a listing of the growing number of evidence-based practice medical and mental health centers in the U.S. and Canada.

² Lowenkamp, Christopher T. and Edward J. Latessa. 2005. "Evaluation of Ohio's CCA Funded Programs. Final Report." Cincinnati, OH: University of Cincinnati, Center for Criminal Justice.

³ Austin, James. 2006. "What Should We Expect From Parole?" American Probation and Parole.

⁴ For a recent summary of the validity of the LSIR and its history see Girard, Lina and J. Stephen Wormith, 2004. "The Predictive Validity of the Level of Service Inventory- Ontario Revision on General and Violent Recidivism among Various Offender Groups" *Criminal Justice and Behavior*, Vol. 31, No.2:150-181 and Violent Recidivism among Various Offender Groups" *Criminal Justice and Behavior*, Vol. 31, No.2:150-181.

⁵ For more information on MHS, Inc see their website at <http://www.mhs.com/index.htm>.

⁶ For more information about Northpointe see their website at <http://www.northpointeinc.com/contact.htm>.

⁷ Washington State Institute for Public Policy. *Washington's Offender Accountability Act: An Analysis of the Department of Corrections' Risk Assessment*. December 2003. Olympia, Washington; James Austin, Dana Coleman, Kelly Dedel-Johnson, and Johnette Payton. 2003. *Reliability and Validity of the LSI-R Risk Assessment Instrument*. Washington, DC: The Institute on Crime, Justice and Corrections at the George Washington University; and James Austin, 2006. *Vermont Parole Board Risk Based Guidelines, Technical Assistance Report #2*. Washington, DC: National Institute of Corrections.

⁸ Washington State Public Policy Institute. 2003. p.4

⁹ Austin, James, Dana Coleman, and Kelly Johnson. 2002. "Reliability and Validity of the LSI-R for the Pennsylvania Board of Parole and Probation." Washington, DC: The Institute on Crime, Justice and Corrections, The George Washington University.

[back to top](#)

From Theory to Practice: The Lifecycle Document for the Results-Based Management Framework for the Federal Probation and Pretrial Services System

¹ The terms "outcome-based" and "results-based" are used interchangeably.

² The term "evidence-based practice" implies that 1) there is a definable outcome(s); 2) it is measurable; and 3) it is defined according to practical realities (e.g. recidivism).

³ As shown by the framework development flow chart, the process is iterative. All references to “completion” refer to the initial development process.

⁴ While the focus of the measurement matrix is on outcomes, examples of measures for the inputs depicted in the logic model have been provided. It is expected that these will be revised and refined based on review and comment from system staff and stakeholders.

⁵ While most of the imputation-based procedures described are available through SPSS, other more advanced procedures may require additional statistical software, such as SAS or Stata.

⁶ For principal component analysis, variance maximizing (varimax) rotation should be selected as the extraction method. Additionally, different criterion (Kaiser criterion, scree test) should be examined to determine which solution makes the best sense, often one retaining more factors (Kaiser) than the other (scree).

⁷ There are over 25 goodness-of-fit calculations available through the SPSS add-on AMOS. The most common used are model chi-square (not significant indicates model fit), GFI (goodness-of-fit index) (greater than or equal to .90 to accept the model), and CFI (comparison-fit-index) (greater than or equal to .90 to accept the model). Any or all of these should be compared.

⁸ Path analysis and structural equation modeling can be conducted using SPSS and the SPSS add-on software AMOS as well as SAS. necessary to modify the model (add and/or remove arrows depicting relationships, add/or remove variables, etc.).

⁹ Most univariate analysis can be conducted using the Descriptive Statistics option within the Analyze function of SPSS.

¹⁰ Bivariate analysis can be conducted by using the Analyze function of SPSS and selecting Descriptive Statistics, Compare Means, and Correlate options.

¹¹ When comparing means using ANOVA, multiple range tests are used, the most popular of which is Tukey’s HSD procedure.

¹² ANOVA can be conducted by using the Analyze function in SPSS and selecting the Compare Means/One-Way ANOVA option.

¹³ This particular statistical technique is not available through SPSS. It would require the use of Stata statistical software.

¹⁴ Survival analysis can be performed using the Analyze function and selecting the Survival/Cox Regression option in SPSS.

¹⁵ While you can use SPSS to generate trend graphs, a software add-on, SPSS Trends, is required to conduct more sophisticated time series analysis.

¹⁶ When comparing across different populations, it is important to use b’s (regression coefficients) rather than β ’s (standardized regression coefficients) because they are more sensitive to fluctuations in variances and covariances across populations.

[back to top](#)

[References](#)

