

NCCD Compares Juvenile Justice Risk Assessment Instruments: A Summary of the OJJDP-Funded Study



Acknowledgments

The National Council on Crime and Delinquency wishes to acknowledge that the study summarized here was funded by Grant 2010-JR-FX-0021 from the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, US Department of Justice. Points of view in the original report and this summary are those of the authors and do not necessarily reflect the official position or policies of the US Department of Justice.



NCCD promotes just and equitable social systems for individuals, families, and communities through research, public policy, and practice.

Introduction



Juvenile justice service staff began to seriously explore the use of actuarial risk assessments in the 1970s, seeking to classify offenders by their likelihood of future delinquency. In 1998, the Office of Juvenile Justice and Delinquency Prevention (OJJDP) made clear the relevance of a valid, reliable, and equitable risk assessment within a broader juvenile justice reform effort when it published *A Comprehensive Strategy for Serious, Violent, and Chronic Juvenile Offenders*. OJJDP's strategy illustrated how juvenile justice agencies could better ensure the effectiveness and appropriate targeting of services by implementing the use of two risk assessments: an actuarial risk assessment to accurately, reliably, and equitably classify youth by the likelihood of future delinquency and an equally effective needs assessment to identify an intervention and treatment plan tailored to each individual.

As other models of risk assessment were introduced, researchers began categorizing and comparing them as generations of risk assessments. The first generation of risk assessments were not actuarial—individual workers assigned risk levels without the aid of actuarial instruments. Generation 2 instruments were statistically derived, but relied heavily on static criminal history factors to assess risk.

Many of today's risk assessment instruments, often referred to as generation 3 or 4, have expanded beyond the original objective of classifying individuals by risk of delinquency. These instruments often contain dozens of risk factors that are divided into two groups: "static" and "dynamic" (see, for example, Schwalbe, 2008; Hoge, 2002). Static factors are generally measures of prior delinquency. Dynamic factors are commonly referred to as "criminogenic



needs” and represent conditions or circumstances that can improve over time (Andrews, Bonta, & Wormith, 2006). In addition, protective factors and references to “responsivity” have been added to generation 4 instruments. Responsivity is intended to reflect an individual’s readiness for change and gauge a youth’s ability to respond to particular treatment methods and programs (Andrews, 1990). Generation 4 instruments contain anywhere from 42 to approximately 150 factors.

These variations in methodology and philosophy have raised questions about the types of instruments that most accurately and effectively help jurisdictions differentiate between low-, moderate-, and high-risk

youth. In addition, few jurisdictions have conducted local validation studies to ensure a risk assessment’s validity and reliability; now one foundation-funded reform effort is telling agencies that local validation is not required if an instrument has been validated in three agencies or for similar populations. The most significant change in the last few decades may be the emergence of commercially available risk assessment systems. Prior to this, risk assessment studies were generally conducted by universities, nonprofit research organizations, or research units within government agencies. Claims made about the validity and reliability of some of these instruments have been challenged by other researchers (Skeem & Eno Louden, 2007; Baird, 2009).

In response to concerns voiced by juvenile justice practitioners and researchers about the classification and predictive validity of several risk assessments, OJJDP funded an evaluation of those most commonly used. The National Council on Crime and Delinquency (NCCD), a nonprofit social research organization, conducted the study of eight risk assessments in 10 jurisdictions in consultation with an advisory board of juvenile justice researchers and developers of commercial juvenile justice risk assessment systems included in the study. The study compared the assessments’ predictive validity, reliability, equity, and costs.

The Study

The 10 jurisdictions use a variety of risk assessment instruments, ranging from commercially available systems to models developed for use by a specific agency. The seven agencies that use risk assessment models created for general use include the Arkansas Department of Human Services, Division of Youth Services; Florida Department of Juvenile Justice; Georgia Department of Juvenile Justice; Nebraska Department of Health and Human Services, Office

of Juvenile Services; Nebraska Office of Probation Administration; Solano County, California; and the Virginia Department of Juvenile Justice. The three assessment instruments used in Arizona and Oregon were validated on and for local populations.

Table 1 lists each assessment that was tested, along with a very brief summary of performance findings. More detail regarding testing follows the table.

Table 1
Comparison of Risk Assessment Instruments in Juvenile Justice

Site Agency	Risk Assessment Instrument	Summary of Findings
Arizona Administrative Office of the Courts (AOC)	Risk/needs assessment for Arizona youth placed/referred to juvenile court	Did not perform well.
Arizona Department of Juvenile Corrections (DJC)	Dynamic Risk Instrument (DRI) for secure care/committed population	Complex formulas produced moderate results.
Arkansas Department of Human Services, Division of Youth Services (DYS)	Youth Level of Service/Case Management Inventory (YLS/CMI) for youth in secure commitment	Did not perform well.
Florida Department of Juvenile Justice (DJJ)	Positive Achievement Change Tool (PACT)	Produced a moderate degree of discrimination.
Georgia Department of Juvenile Justice (DJJ)	Comprehensive Risk/Needs (CRN) assessment	Large number of factors and complex scoring system did not appear to help CRN produce better results.
Nebraska Department of Health and Human Services, Office of Juvenile Services (OJS)	YLS/CMI for youth in secure commitment	Did not perform well.
Nebraska Office of Probation Administration	YLS/CMI	Did not perform well.
Oregon Juvenile Justice	Juvenile Crime Prevention (JCP) assessment for youth referred to juvenile justice system	Due to low rates of recidivism in Oregon, results from JCP are difficult to compare to other assessment instruments studied.
Solano County [California] Probation Department	Juvenile Sanctions Center (JSC) and Girls Link (gender-specific risk assessments in JAIS for youth referred to probation)	JSC was found to be an effective classification instrument. Girls Link was effective, but not at the same rates as JSC. Both worked well across major racial and ethnic groups in Solano County.
Virginia Department of Juvenile Justice (DJJ)	Youth Assessment and Screening Instrument (YASI) for youth on probation or parole and in facilities	Worked well overall; better for boys than for girls.

Inter-Rater Reliability Testing

Inter-rater reliability is a necessary quality in an assessment because it helps ensure that different caseworkers, faced with the same case information, will reach the same scoring and recommendations for key decision thresholds such as risk of future delinquency. NCCD measured the inter-rater reliability of risk assessment items by asking a sample of officers/caseworkers to review case files for 10 youth, observe a videotaped interview of each youth, and score a risk assessment (or risk/needs assessment) for each youth. Multiple measures were used to assess inter-rater reliability, as each has limitations that are important to understand. Percent agreement is NCCD's primary measure for comparison across items and assessments because it is easy to understand; the limitation is that it does not control for the likelihood of caseworkers randomly reaching the same responses by chance.

In a comparison of assigned risk level by each assessment for 10 test cases, most instruments achieved high percent agreement between workers. Fewer instruments achieved high levels of agreement with an expert score (five of the 10). Of most interest is that only three of the risk assessments had positive indications of inter-rater reliability across every measure: Arizona's homegrown AOC assessment, Solano County's gender-specific assessments, and Virginia's YASI. Overall, prior delinquency history and other similar static risk factors demonstrated higher levels of inter-rater agreement than dynamic factors; this was especially true for more subjective measures such as youth attitudes.

Validity and Equity Testing

In order to effectively target limited resources, a risk assessment needs to result in valid and equitable classifications. Testing the predictive validity and equity of the risk assessments involved sampling a cohort of youth on probation or released from a facility. Recidivism was tracked over a 12-month follow-up period for all sites but one, where only nine months of outcomes were available. Outcome

measures were obtained from agency databases and included subsequent arrests, subsequent adjudications, and subsequent juvenile facility placement. Exceptions were two sites for which recidivism was limited to return to a correctional facility for youth released from facilities. In these two sites, data on new arrests and adjudications were not available. Findings showed that several of the evaluated risk assessment systems failed to provide the level of discrimination needed by probation and correctional service staff if they are to optimize decisions regarding supervision requirements.

Three systems—the Oregon JCP, Solano County's JSC risk assessment for boys, and the YASI model used in Virginia—demonstrated considerable capacity to accurately separate cases into low, moderate, and high risk levels with progressively higher recidivism with each risk level increase. Of these three systems, the YASI is the most complex; the JSC is the easiest to complete. Unfortunately, YASI results were available for only a small segment of the Virginia population. When this study began, YASI was being implemented statewide, but only about 20% of the state's cases had been assessed. All other systems evaluated were fully implemented well before the study was initiated.

Potential equity problems were found with the YASI. It did not work well for females, and moderate-risk



African American males were only slightly less likely to recidivate than high-risk White males. The Oregon assessment proved valid for girls; in Solano County, a separate risk instrument is used for girls. These two approaches worked well across all ethnic/racial groups tested. No other instrument tested in this study provided the level of discrimination needed to support decision making in the juvenile justice system.

In all jurisdictions where sample size allowed, NCCD conducted additional analyses to determine if a simple actuarial risk instrument would provide better classification results. This effort was restricted by available data, but better results were obtained in most instances using simple construction scale methods such as analyses of correlations and regression models. In two agencies with large study cohorts available, cases were divided into construction and validation samples and results from the validation samples presented. This step is recommended because results from a construction are generally the best that will be attained. When tested on an independent sample, the level of discrimination attained tends to decline. In this exercise, NCCD found minimal “shrinkage.” The combined results of all analyses conducted suggest that limiting factors on a risk assessment to those with a strong, significant relationship to outcomes will result in a more accurate risk classification.

In short, risk assessments should be evaluated based on how the information informs practice; thus, NCCD assessed predictive validity using multiple measures, with recurrence of delinquency by risk classification level as the primary measure.

Implications for Practice

The proper use of valid, reliable risk assessments can clearly improve decision making. Results of this study show, however, that the power of some risk assessment instruments to accurately classify offenders by risk level may have been overestimated. The first step in remedying this situation is to ensure

that everyone working in the field of juvenile justice understands the importance of valid, reliable, and equitable risk and needs information. Although the study provided fodder for many areas of policy and practice, as well as future research and development, researchers, practitioners, and advocates should focus attention on the following points.

- Jurisdictions must be able to ensure that the risk assessment completed by field staff to inform case decision making is reliable, valid, and equitable. Decisions about youth are based on the level of risk assigned. Thus, the primary measure of validity must be the level of discrimination produced. Jurisdictions should expect reliability testing and validation studies when assessment models are transferred to other jurisdictions; they would benefit from making assessment evaluation part of a more comprehensive approach to evidence-based practice.
- National standards could provide juvenile justice administrators with clear guidelines for assessing the reliability, validity, and equity of existing models. Such standards could also help agencies develop the capacity to construct instruments for their populations and understand how valid risk and needs information can help them monitor and improve practice.
- Risk assessment should focus solely on identifying cases most and least likely to be involved in future offending.

Risk assessment should be a simple process that is easily understood and articulated. This study's findings show that simple, straightforward, actuarial approaches to risk assessment can produce the strongest results.

References

- 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(1), 7.
- Andrews, J. D. (1990). Interpersonal self-confirmation and challenge in psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, 27(4), 485.
- Baird, C., Healy, T., Johnson, K., Bogie, A., Dankert, E. W., & Scharenbroch, C. (2013). *A comparison of risk assessment instruments in juvenile justice*. Madison, WI: National Council on Crime and Delinquency. Baird, C. (2009). Special Report.
- Hoge, R. D. (2002). Standardized instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior*, 29, 380–396.
- Schwalbe, C. S. (2008). Strengthening the integration of actuarial risk assessment with clinical judgment in an evidence based practice framework. *Children and Youth Services Review*, 30, 1458–1464.
- Skeem, J., & Eno Louden, J. (2007). *Assessment of evidence on the quality of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)*. Unpublished report prepared for the California Department of Corrections and Rehabilitation. Available at www.cdcr.ca.gov/Adult_Research_Branch/Research_Documents/COMPAS_Skeem_EnoLouden_Dec_2007.pdf