Summary and Recommendations

Chris Baird, Chair, NCCD Board of Directors
In this series of briefs, I described problems with risk assessment models that have emerged over the last two or three decades. The sources of these problems are varied, ranging from poorly designed research, flawed logic, and misrepresentations of older, well-established risk assessment systems to the proliferation of for-profit vendors that sell and support risk assessment models. To summarize, the major issues identified include the following.

1. Most newer systems are not truly actuarial. These systems were frequently marketed before being adequately tested for validity, reliability, or equity.

2. Language associated with the goals and objectives of risk assessment has changed significantly, suggesting a level of precision that far exceeds what can be legitimately inferred from available research.

3. For many years, the principal measures of validity, the degree of discrimination attained between observed outcomes for cases assigned to different risk levels, were largely ignored. Criticisms have had some effect and, lately, there has been something of a return to standard measures of validity.

4. In promoting the use of new “generation 3” and “generation 4” risk assessment models, developers misrepresented existing models and ignored important research conducted for county, state, and federal agencies. Conclusions that the new systems offered greater “predictive validity” and better reflect changes over time have been thoroughly refuted.

5. The emergence of the for-profit sector in the development of risk assessment systems represents a major change in the justice and corrections landscape. Historically, risk assessment research was conducted by universities, nonprofit research organizations, and state research offices, often funded by grants from the federal government. Risk instruments were generally developed for individual agencies, reflecting the laws, policies, and populations of each jurisdiction. Furthermore, because they were developed with public funding, most systems were in the public domain. Today, many in the research community promote risk assessment models that are “transferable” or “generalizable.” There are two major problems

NCCD | National Council on Crime & Delinquency
with this approach. First, as noted above, many models were marketed before being adequately tested for validity or reliability. Second, even if valid and reliable, these instruments will not perform optimally in all jurisdictions as they fail to reflect local policy, law, practice, and population differences.

The ability to purchase a validated risk model can be an attractive alternative to the time, cost, and effort associated with system development, and could be more effective if developers encouraged agencies to make changes based on follow-up research in each jurisdiction. However, firms know that supporting various renditions of a model is difficult (and therefore costly), so customization is not always encouraged.

These developments come at considerable cost to the efficacy of decision making. In many instances, follow-up research indicates these systems do not work as intended (Flores, Travis, & Latessa, 2004; Baird et al., 2013). But these findings seem to have little impact, as marketing continues without revisions to existing models. Such results are often ignored by the research community and attributed to lack of fidelity to the model.

While these briefs paint a bleak picture of current practice, remedies need be neither difficult nor costly. Risk systems can be easily streamlined and improved, given the amount of data now available in most agencies. However, to stem the proliferation of instruments that fail to optimally discriminate between high-, moderate-, and low-risk offenders, both researchers and correctional administrators must be clear about the objectives of risk and needs assessments. The following steps would improve assessment practice and add the clarity that is desperately needed if assessment is to optimally guide decision making in corrections.

1. First, justice agency officials must better understand the roles of risk and needs assessments. Marketing strategies are often cloaked in research and statistical terminology, and few administrators have the technical background needed to effectively evaluate claims of validity and reliability. Having research expertise on staff would be valuable, but smaller agencies seldom have this “luxury.” As an alternative, federal agencies such as the National Institute of Corrections, the National Institute of Justice, and the Office of Juvenile Justice and Delinquency Prevention could establish guidelines for assessment practice, including guidance on interpreting research results. This could significantly increase understanding at the agency level and improve both decision making and outcomes. These guidelines are sorely needed. Vendor claims of “predictive accuracy” will be more difficult to evaluate as statistical modeling becomes more complex and larger firms, with greater marketing potential, enter the field.

2. The research community must become far more self-regulating. Peer review needs to improve, and uncritical acceptance of ideas published in prior articles should end. Although beyond the
scope of this paper, the volume of error found in journal articles and other publications is astounding.\(^1\) Critiques should play an essential role in knowledge development and transfer. Recently, however, criticizing accepted views has opened researchers to attacks from both developers and users,\(^2\) which discourages challenges to widely accepted assessment and treatment protocols. As a result, the justice field has been dominated by a few individuals whose work has attained a status that few researchers are willing to challenge. At the same time, excellent publications by other respected researchers have been basically ignored, rarely cited by vendors marketing specific models (e.g., Gottfredson & Snyder, 2005; Gottfredson & Moriarty, 2006; Austin, Coleman, Peyton, & Johnson, 2003; and Fabelo, Nagy, & Prins, 2011).

3. The inconsistencies, flawed logic, and inadequate research that have permeated the field over the last two decades must be addressed. In this era of evidence-based practice, justice and corrections officials rely on the research community to provide guidance in selecting both assessment and treatment programs. When issues of professional advancement and financial interest enter the equation, the potential for abuse increases. While these issues will never be eliminated, everyone needs to understand they exist. These briefs identify problems with the research that supports many widely used approaches to assessment, flaws in the design of many models, and statements meant to provide guidance that are simply inaccurate. Given the degree to which emphasis on evidence-based practice influences policy and practice, it is essential that the evidence provided is accurate, unbiased, and open to critique.

4. Finally, proponents of new analytical methods (e.g., neuro-networking and random forest analysis) are making unprecedented claims of predictive accuracy. These claims should

---

\(^1\) Many statements are made in journals and reports without citation or evidence presented to substantiate their content. In other instances, the evidence presented does not support the claims made. Unfortunately, many such statements are repeated in subsequent publications. Several such statements are presented in this series.

\(^2\) For example, one researcher’s well-constructed critique of research used to support multisystemic therapy was met with a deluge of unwarranted criticism from the industry (Littell, 2006).
be viewed with caution and fully vetted by the research community. Often, the results are no better than those produced by well-designed studies that use traditional methods of scale development. Other fields—medicine, for example—also have found that the new analytical techniques provide no benefits over standard scale development methods (Altman & Royston, 2000). These methods sometimes produce highly suspect results, overstated by omitting large segments of the target population, and can result in predictive models that have little decision-making utility. Furthermore, developers of these models are often reluctant to share their algorithms with users, claiming they are “intellectual property.” As a result, these systems become “black box” models, as users are not provided with the criteria used to assign risk levels. There are two major issues with black box systems. First, workers cannot determine if an override is appropriate because the basis for assignment to a risk level is unknown. Second, no one can effectively challenge any decisions influenced by risk assessment, again because information used to make the decision is unknown. When these tools are used in sentencing or release decisions, their constitutionality can and should be challenged. As prior briefs have noted, many existing models contain factors that have no relationship to recidivism or to violence. In addition, black box models could well contain factors that discriminate against specific groups of people (e.g., people of color), or factors that should have no role in sentencing, placement, or parole decision making. Transparency should be a requirement, and no jurisdiction should allow black box systems any role in decision making.
References


