
Jennifer Skeem

The target article is a critique of the movement toward using structured risk assessment tools to inform decisions about sentencing. In this commentary, I analyze (a) the conditions under which it may be more or less fair to use well-validated risk assessment tools in this manner and (b) the extent to which doing so is likely to exacerbate, ameliorate, or have no effect on existing racial and other biases in sentencing. I recommend a policy-relevant research agenda that would specifically test whether and how adding well-validated risk assessment tools to the routine sentencing process alter the severity or nature of sentences. This agenda would also evaluate the extent to which these tools are implemented in “real world” settings faithfully enough to bridge the usual divide between science and practice.

Keywords risk assessment; sentencing; violence; crime; actuarial

Few ideals have gained greater traction in contemporary discourse than “evidence-based practice.” According to this ideal, the best research informs

Jennifer Skeem is a member of the MacArthur Research Network on Mandated Community Treatment, and the Centers for Psychology and Law and Evidence-Based Corrections at UCIrvine. She earned her PhD from the University of Utah and completed postdoctoral training at the University of Pittsburgh Medical Center. Her research is designed to inform clinical and legal decision-making about individuals with mental disorder. She has authored and coauthored over 70 articles, chapters, and books. To help research inform policy and practice, she works closely with both national and local agencies. Dr. Skeem has received the Saleem Shah Award for Early Career Excellence from the American Psychological Association (Division 41) and the Distinguished Assistant Professor Award for Research from the Academic Senate of UCIrvine. Correspondence to: Jennifer Skeem, Professor of Psychology & Social Behavior, University of California, Irvine, CA, USA. E-mail: skeem@uci.edu
practice that improves outcomes. Two bodies of research have helped fuel a recent movement toward evidence-based sentencing and corrections. First, research has established that using validated, structured tools (like the LSI-R) significantly improves professionals’ ability to predict future violent or other criminal behavior (see Skeem & Monahan, 2011). Increasingly, these tools are being applied in response to statutes and regulations that require specialized assessments to identify "high risk" individuals for detention or "low risk" individuals for release. Second, research suggests that correctional programs reduce recidivism when they (a) match the intensity of services and supervision to an offender’s level of risk and (b) target robust risk factors for crime (e.g. criminal attitudes) rather than variables that are less crime relevant (e.g. low self-esteem; see Andrews, in press). Increasingly, companies are marketing tools for corrections agencies that ostensibly include changeable risk factors and inform risk reduction efforts.

The time is ripe for a critical review of this movement. Hannah-Moffat (2011) distills the promises and perils of using risk assessment technology in sentencing and corrections. Although she believes that using this technology could reduce bloated prison populations by diverting low-risk offenders and promoting effective intervention, she raises pointed concerns that doing so (1) is not fair because it may "punish individuals for crimes they have not committed" and "undercut proportionality" in sentencing (p. 18) and (2) will exacerbate existing bias against marginalized groups in the criminal justice system (e.g. racial minorities).

In this commentary, I analyze these two concerns before offering one of my own. My goal is to stimulate discussion and research that sheds light on the actual good and ill effects of using risk assessment tools in sentencing and corrections. It is important to state two premises before I begin. First, in many fields of risk assessment (from determining insurance premiums to forecasting the weather), group data often are highly informative when making decisions about an individual case (Skeem & Monahan, 2011). Even though validated risk assessment tools cannot predict an individual’s violent or other criminal behavior with certainty, there is compelling evidence that structuring or even replacing professional judgment with these tools substantially improves predictive accuracy. Second, the value of using a risk assessment tool to inform sentencing depends on (a) the strength the tool’s evidence base and implementation in a given setting, (b) whether it is appropriate to weigh crime control goals in reaching a given decision, and (c) how the process and outcome of decisions that are informed by the tool compare to traditional/unstructured decisions.

When is it Fair to Use Well-Validated Risk Assessment Tools in Sentencing?

Although traditional sentencing ostensibly is a backward-looking process that focuses on blameworthiness for past conduct, risk assessment tools necessarily
look forward to the likelihood of a future crime, based on past conduct, personality traits, substance abuse, and/or other risk factors (Monahan, 2006). Science aside, it seems unjust to assign punishment based “not on what offenders did, but rather on how closely who they ‘are’ approximates subgroups of an offender population” (Hannah-Moffat, 2011, p. 6).

Given such concerns, Monahan (2006) argues that “the use of violence risk factors in sentencing is jurisprudentially constrained to those that index the extent or seriousness of the defendant’s prior criminal conduct.” He reasons that concerns for just deserts are strong enough in sentencing that they should constrain the variables used in pursuit of crime control. In contrast, he argues that the use of violence risk factors to make decisions about civil commitment is unconstrained, with the exception of race (given equal protection concerns). This is because civil commitment is designed not for punishment, but for protection of public safety. Reasoning from Monahan’s analysis, no well-validated risk assessment tool that I am aware of would qualify for use in sentencing, but virtually all of them would qualify for use in civil commitment.

Although this bright line analysis is helpful, it does not resolve shades of gray that arise under four specific conditions. First, when civil commitment strays from detaining an offender to prevent further crime to detaining an offender as further punishment for past crime, it seems inappropriate to rely upon risk assessment technology. Increasingly, countries are enacting preventive detention laws that allow for the indefinite commitment of offenders who may remain dangerous after having served their period of punishment (e.g. sex offenders). By default, these offenders’ criminal history (including crimes they have already been punished for) will be weighed heavily by risk assessment tools. Moreover, there is evidence that support for the civil commitment of sex offenders is driven more by just deserts concerns than crime control goals (Carlsmith, Monahan, & Evans, 2007).

Second, when sentencing cares about crime reduction, it may be appropriate to rely upon risk assessment technology to inform these efforts. An alternative to discarding risk assessment tools from sentencing altogether is to explicitly tether them to the crime control issue. For example, a two-stage sentencing process might be entertained. In Stage 1, a term of incarceration or community supervision would be imposed based on just deserts principles and past criminal conduct. In Stage 2, crime reduction principles and well-validated risk assessment tools would be used to mandate supervision and interventions that (a) are matched in their intensity to the offenders’ level of risk (such that lower risk offenders have fewer requirements) and (b) are targeted to offenders’ causal risk factors for crime. As an alternative example, risk assessment tools could be used at Stage 2 to divert low-risk offenders from prison into alternative sanction programs (see Kleiman, Ostrom, & Cheeseman, 2007). This application directly addresses Hannah-Moffat’s (2011) concern that using this technology may over-penalize offenders.

Third, when sentencing overlaps with correctional programming, it may be appropriate to rely upon risk assessment technology. Sentencing sometimes
overlaps with corrections. For example, (a) judges leave some conditions of supervision to the discretion of a probation agency (e.g. participation in specific treatments), (b) the conditions of supervision can be changed during the course of a sentence, and (c) an offender can earn parole (release from incarceration) or may have a term of community supervision revoked. If these decisions were construed as part of sentencing, then well-validated risk assessment tools could not be used to inform them. However, the backward-looking focus must be loosened at some point, if the system is interested in behavior change. For example, a lifer with the possibility of parole can be told to develop and demonstrate more prosocial attitudes, but will never be able to undo his or her index offense.

Fourth, when measured against the yardstick of existing practice rather than theory, the use of risk assessment technology to inform sentencing may appear quite appropriate. Arguably, the benchmark is not what ought to be considered in sentencing, but instead what actually is considered. From a practical point of view, the “compared to what?” question for risk-assessment-based sentencing is existing practice. Typically, judge’s sentencing decisions are driven by presentence investigation reports (PSIs) completed by probation agents. Some of these reports may focus narrowly on the defendant’s criminal record, but most will include information that goes well beyond that (e.g. age, living arrangement, family, employment, sophistication, remorse, and related traits). Many reports will speculate about the offender’s risk for recidivism before recommending a sentence. There is probably considerable variation across jurisdictions in the extent to which a probation officer (and, in turn, a judge) fits the sentence to the crime or to the offender. Thus, it is possible that standardizing the risk assessment portion of PSIs with a well-validated tool often improves the transparency, consistency, and fairness of decision-making. It is equally possible that doing so has no effect on sentences. Because criminal history variables are strongly correlated with other leading risk factors for crime (e.g. an irresponsible lifestyle, antagonistic, or impulsive traits), there may be little difference between sentences based on criminal history alone vs. those that include other risk factors. For these reasons, rigorous experiments (with real or hypothetical cases) are needed to determine whether and how adding well-validated risk assessment tools to the sentencing process alters (a) the weighting of criminal history vs. other factors and (b) the severity or nature of sentences.

Does Using Well-Validated Risk Assessment Tools Exacerbate Existing Sentencing Biases?

Hannah-Moffatt’s (2011) second major concern is that introducing risk assessment technology to sentencing will exacerbate existing bias against marginalized groups. The simplest and most direct way to address this concern is by conducting the experiments recommended above. Indeed, group-based data are ideal for revealing bias that can lurk undetected within individual cases,
whether they are or are not informed by technology. In the remainder of this section, I address more specific aspects of this general concern.

Hannah-Moffat (2011) worries that risk assessment tools have been developed in a manner that predicts not crime per se, but biased criminal justice decisions. All too often, studies rely solely upon official records to measure the criterion of violence or other criminal behavior. However, studies that have relied upon different sources of information (e.g. victimization surveys and self-report) converge on major risk factors for violence and crime, including such controversial variables as race (see Monahan, 2006). This mitigates concern that these variables predict biased arrest/prosecution practices rather than criminal behavior per se.

Hannah-Moffat’s (2011) broader worry is that risk assessment tools individualize social problems like racial discrimination and poverty. Similarly, Harcourt (2007) argues that the variable heavily emphasized in risk assessment tools—criminal history—is a proxy for being black. It is not clear that this is the case. For example, Andrews (in press) has demonstrated that well-validated risk assessment tools predict recidivism well above and beyond the effects of race, poverty, and gender. This casts doubt on the notion that robust individual variables (e.g. angry, impulsive traits and criminal behavior) are nothing more than masked indices of social disadvantage. Additional doubt is cast by research indicating that the predictive utility of well-validated risk assessment factors and tools often generalize across race, gender, and even age.

I suspect that Hannah-Moffat (2011) will be unimpressed by evidence that risk assessment measures generalize across disadvantaged groups. Instead, she seems interested in the generalizability of the constructs tapped by these measures—including their underlying mechanisms and causes. There is evidence that well-validated tools assess constructs like chronic criminal behavior, an irresponsible lifestyle, psychopathy and criminal attitudes, and substance abuse-related problems (Kroner, Mills, & Reddon, 2005). It is likely that many of these constructs are ultimately caused by intricate interactions among many kinds of biologic and environmental variables. Given how little is known about ultimate causes and how unlikely those causes are to sort neatly into “blameworthy” and “non-blameworthy” categories, I recommend that we first focus on the generalizability of risk factors immediately relevant to risk assessment and risk reduction.

In shaping a policy-relevant research agenda on this issue, two considerations are important. First, the factors that are central to risk assessment and risk reduction are not necessarily equivalent. For example, mental illness is a risk factor for criminal behavior, but there is no compelling evidence that reducing symptoms will reduce recidivism (Skeem, Manchak, & Peterson, 2011). Second, there are differences between risk factors and causal risk factors (see Kraemer et al., 1997). We know a great deal about risk factors, i.e. variables that precede and increase the likelihood of crime. This puts us on relatively solid footing for risk assessment. We know less about causal risk fac-
tors, i.e. risk factors that reduce the likelihood of crime when successfully changed in treatment (see Douglas & Skeem, 2005). Even well-validated tools offer little direct validity data for risk factors that ostensibly are causal (see Skeem & Monahan, 2011). To inform risk reduction efforts, we need to identify robust causal risk factors and assess their generalizability across race, gender, and age.

Are Risk Assessment Tools Implemented Rigorously Enough in Practice to be Informative?

Having addressed Hannah-Moffat’s (2011) main concerns, I add an overlapping concern of my own. The most robust risk factors for crime include a history of criminal behavior and personality traits like impulsivity and anger or antagonism (e.g. Monahan, 2006; Skeem, Miller, Mulvey, Monahan, & Tiemann, 2005). These risk factors are common to virtually all well-validated risk assessment tools, despite variation in risk factors and the degree of structure with which they are assessed (see Skeem & Monahan, 2011). Perhaps for this reason, there is no compelling evidence that one well-validated instrument predicts much better than another (Yang, Wong, & Coid, 2010).

One might conclude that we have hit a ceiling with predictive accuracy and should shift attention to identifying causal risk factors that can be targeted in risk reduction efforts. However, the priorities for scientific understanding and practice may be quite different. It could be that the risk assessment tools applied in “real” settings do not even improve our ability to predict criminal behavior. First, increasingly complex and poorly validated risk assessment tools are being sold to criminal justice agencies (see Baird, 2009; Skeem & Eno Louden, 2007). Second, even for well-validated tools, implementation efforts can fall breathtakingly short.

Research is needed to describe the “lay of the land” for practices in risk assessment. What risk assessment tools are being applied by criminal justice agencies? How well-validated are they, in the abstract? What training do staff receive on instrument? Can the agency document that staff (a) attain and maintain interrater reliability with a scoring criterion (because a tool cannot be valid if it is not consistently scored) and (b) produce scores that correlate with theoretically related measures or predict violent or criminal behavior? My experiences are consistent with Hannah-Moffat’s (2011) observations that staff can assign scores that are based on incomplete data or limited sources of information, that are deliberately “tinkered with” to manipulate programming, and that reflect individual biases. These observations resonate with direct evidence that even forensic experts “tinker with” scores on risk assessment tools (Murrie, Henderson, Vincent, Rockett, & Mundt, 2009). More troubling is our overlapping observation that agencies pay lip service to using risk assessment to inform programming, but do not actually attach services to the technology.
Risk reduction will be achieved through risk assessment alone, even if a well-validated tool is successfully implemented. In short, we must know more about how the promise of risk assessment technology is being realized in practice before we determine whether it is appropriate to rely on this technology to inform sentencing.

References


