Decision Points for Risk Assessment Implementation

Megan E. Collins and James P. Lynch[1]
Department of Criminology and Criminal Justice
University of Maryland

This brief document is designed to assist the Maryland State Commission on Criminal Sentencing Policy (MSCCSP) and the Judiciary's Risk Assessment Advisory Group (JRAAG) in making decisions regarding whether and how to implement a risk assessment tool that could be used at sentencing. Staff from the University of Maryland reviewed the academic literature on risk assessment tools as well as the experience of other states in building and using these tools. On the basis of this review, we identified major decisions that must be made to implement a risk assessment tool, provided recommendations regarding these decisions and offered justifications for those recommendations. Many of the decisions presented here pertain to the Commission's goals in using the tool, since decisions on these issues will affect choices pertaining to the structure of the tool. Other decisions pertain directly to how the tool will be developed and implemented. We ask the MSCCSP and the JRAAG to respond to the recommendations and justifications made here and to add any important decisions about the design of the tool not addressed in this paper.

1. Should actuarial tools be used to predict future recidivism at sentencing, instead of or in conjunction with professional judgment?

Recommendation:

If public safety is accepted as a major concern in sentencing, then judges should use actuarial tools to predict offender re-offending.

Justification:

There are strengths and weaknesses to both clinical judgment and actuarial approaches to sentencing and risk assessment. To be clear, actuarial assessment refers to a formal methodology that provides "a probability, or expected value, of some outcome. It uses empirical research to relate numerical predictor variables to numerical outcomes. The *sine qua non* of actuarial assessment involves using an objective, mechanistic, reproducible combination of predictive factors, selected and validated through empirical research, against known outcomes that have also been quantified" (Heilbrun, 2009: 133).

Extensive research comparing human judgment to algorithms has time and again demonstrated that when appropriately calibrated, actuarial models are consistently better at identifying future offenders. There are numerous examples of forecast methodologies applying simple statistical procedures that

[11] Professors Brian Johnson, Thomas Loughran and Kiminori Nakamura are members of an advisory board that reviewed and commented on drafts of this decision memorandum and the literature review on which it is based.

consistently outperform clinical judgment across the criminal justice system (e.g. Skeem & Monahan, 2011; Pew Center for the States, 2011; Casey et al., 2011). Actuarial risk assessment instruments may also be appealing to Maryland residents, as they provide for more transparency, are clearly objective, and are subjected to scientifically rigorous validation procedures (Hyatt et al., 2010; VanNostrand & Lowenkamp, 2013).

On the other hand, some argue that actuarial assessments may be overly rigid, and do not allow for the use of professional judgments in the sentencing decision. They advocate the use of *adjusted actuarial assessments*, which permit the addition of supplemental information not included in the instrument (Austin, 2004). Thus, it would be most prudent to rely on a well validated actuarial tool to calculate explicit, objective risk scores, and rely on experience and clinical judgment to determine how best to utilize the score.[2] Such a process would preserve valuable judicial input, and through the combined information from the actuarial and clinical assessments, the risk of false negatives (i.e. diverting offenders who will eventually reoffend) should be further reduced.

In the past and again more recently, questions have been raised about the appropriateness of using risk assessment tools and about the appropriateness of evidence based sentencing more generally (Starr, 2014). Some object to the use of ascriptive or static characteristics of offenders, such as gender or age, in risk tools because—while these variables can increase the predictive validity of the model—persons should be sentenced for what they have done and not who they are. More achieved statuses like criminal history or employment history are less objectionable—though some argue they can still be problematic, as some ascriptive characteristics may be closely linked to criminal history and employment. While there is not complete consensus over the type of information that can be used in risk assessment tools, this criticism has less to do with whether risk assessment tools should be used, than it does with what information should be used in them.

A second criticism takes issue with the use of patterns of behavior in populations to make inference to particular individuals (Starr, 2014). Specifically, the issue is that risk assessment tools are based upon the offending behavior of large groups of offenders, which is then used to make probabilistic statements about the behavior of individual offenders (economics and sociology literature refers to this phenomena as "statistical discrimination"). There is variance around those estimates, so a specific offender may be a deviant case. He may have all of the attributes that make him likely to offend but he ultimately does not offend. Similarly, some have noted that the margins of error surrounding assessments of individual risk are so wide, that the predictions are of little value (Hart et al., 2007; Cook & Michie, 2010). There are two types of variance in this form of prediction: confidence intervals refer to variance involved in identifying the mean (e.g. the mean level of risk for reoffending); and prediction intervals try to predict individual outcomes from group averages. The variance from these two intervals behaves differently, as one attempts to identify a single measure to represent group behavior, whereas the latter is used to

[2] A validated tool, which will be referenced throughout this document, refers to one that has been empirically tested in order to obtain an unbiased estimate of the instrument's accuracy. This is often done by using half of the sample to construct the instrument, and then using the remainder of the sample to estimate the tool's predictive accuracy (Gottfredson & Moriarty, 2006).

approximate individual behavior using group measures. Because the outcome of the proposed risk assessment instrument is binary (confinement v. non-confinement) the amount of variance is of less concern than the presence of false positives or false negatives.

In its most extreme form, i.e. no estimates from groups can be applied to individuals, this criticism of risk assessment tools is fatal and would lead to the decision not to use them. The force of this critique is blunted somewhat by the fact that the court seems to be willing to make inference from patterns of behavior in groups to the behavior of individuals all the time. This criticism of risk assessment tools can be leveled at any scientific study, yet those studies are used in courts on a daily basis. It is not clear why those inferences are permissible but those made with risk assessment tools are not. No standards are offered for the rate of false negative classifications that is permissible before it can be used in the sentencing decision. Absent the information provided in tools, judges are able to make inferences from their experience or from studies to the offender in unknown ways. Tools establish these patterns in a transparent way and apply them to offenders consistently (Milgram, 2014). It is not that removing risk assessment tools will eliminate inferring individual behavior from patterns observed in groups, but that it will be done in less transparent and less systematic ways.

2. What specifically should the instrument accomplish?

Recommendation:

The instrument should work to divert otherwise incarceration-bound nonviolent offenders to community-based alternatives, without compromising public safety.

Justification:

This goal was identified and agreed upon by the Judiciary Ad Hoc Committee on Sentencing Alternatives, Reentry, and Best Practices (AHCS) and the Maryland State Commission on Criminal Sentencing Policy (MSCCSP). This goal was discussed following the completion of Phase 1 of the three phase approach to implementing risk assessment in Maryland (MSCCSP 2011 Annual Report: pp. 14-15).

3. Who should be targeted?

Recommendation:

The initial instrument should be used to identify low-risk, nonviolent offenders, who would make suitable candidates for non-custodial sanctions.

Justification:

Based on the input of the Maryland State Commission on Criminal Sentencing Policy (MSCCSP) in earlier discussions, including the September 2010 meeting of the Guidelines Subcommittee, it seems clear that stakeholders are invested in using the risk assessment tool to divert nonviolent, low-risk offenders. For example, Dr. David Soulé noted that other risk assessment instruments have been effective in identifying

offenders who are unlikely to recidivate and would make good candidates for community-based sanctions. A subsequent meeting in May 2011 reiterated this point. The Guidelines Subcommittee agreed that it favored using the instrument to target nonviolent offenders. For example, Senator Delores Kelley stated that this approach would be consistent with the MSCCSP's legislative mandate to develop guidelines for identifying defendants qualified for corrections options programs; Dr. Charles Wellford noted that this approach would provide the opportunity to improve upon the implementation and evaluations carried out in Virginia and Missouri. What the Subcommittee means by "nonviolent offender" is less clear however; for example "nonviolent offender" may refer to the individual's current offense, or their entire history; additionally, there is some debate regarding whether some infractions such as burglary or drug sale should be considered nonviolent. Who this group consists of will need to be established before a policy is set in place.

Academic literature supports the notion of targeting nonviolent offenders, as it is generally agreed that high-risk offenders cannot reliably be predicted, whereas low-risk offenders can more consistently be prospectively identified (Bushway & Smith, 2007; Visher, 1986; Gottfredson & Gottfredson, 1986). Bushway and Smith (2007) state that part of the difficulty in identifying high-risk offenders, is that they may already be receiving intensive treatment, and as a result their high-risk behavior may be muted.

4. Should a risk assessment or risk-needs assessment instrument be implemented?

Recommendation:

Short term interests support the full implementation of a risk-assessment instrument (a tool to quantify individual levels of risk); the identification, verification, and evaluation of a *risk-needs* assessment (a tool to assign treatment based on risk scores) should only be considered following the successful launch of a risk-assessment instrument (which will be a sizable task in itself).

Justification:

There are two primary justifications for postponing the implementation of a risk-needs assessment, and focusing on an exclusively risk-oriented tool at the outset: (1) to limit time, resources, and evaluation complexity, and (2) because the tool will be targeting low-risk offenders, who are less likely to recidivate, and generally less likely to benefit from correctional programs than high-risk offenders.

The first justification requires little clarification, and was supported by members of the Guidelines Subcommittee during a May 2011 meeting. Dr. David Soulé invoked the experiences of the Pennsylvania Commission on Sentencing, which initially intended to develop a risk-needs assessment instrument, but ultimately eliminated the needs component due to difficulties encountered; additionally Senator Delores Kelley suggested that the Commission initially focus on developing the risk-assessment instrument, and only revisit the issue of developing a needs instrument once the risk assessment tool is operational. Moreover, reviews of risk/needs assessment tools currently in use indicate that they require substantial amounts of data collection and staff time to implement.

The second justification is based on the empirically supported "risk principle", and related to the earlier recommendation that the instrument target low-risk offenders. The risk principle states that for the greatest impact on recidivism, the majority of services and interventions should be directed toward individuals with higher risk scores (i.e. those with a higher probability of reoffending). This is consistent with the idea of utilizing risk assessment to divert low-risk individuals from custodial sentences, while high-risk individuals are incarcerated, and potentially more likely to receive some form of treatment or programming. Prior research has demonstrated that many interventions are more effective when targeting high-risk offenders' criminogenic needs than those of lower-risk offenders (Bushway & Smith, 2007; Cullen & Gendreau, 2000). Explanations for this trend include that: low-risk offenders are less likely to recidivate and therefore unlikely to benefit from the programming; intensive programing or supervision can potentially interrupt self-correcting behaviors already in place in low-risk offenders; treatment programs may increase low-risk offenders' exposure to high-risk offenders with pro-criminal attitudes; and treatment may disrupt pro-social networks and supports (Latessa, 2004; Casey et al., 2011; Bushway & Smith, 2007; Cullen & Gendreau, 2000). Lowenkamp and Latessa (2004) have identified several meta-analyses supporting the risk principle; additionally, their own research tracked over 13,000 offenders in fifty-three community-based correctional treatment facilities and found that the majority of programs were associated with increased recidivism for low-risk offenders and decreased recidivism for high-risk offenders.

Ultimately, if a *risk-needs* assessment is eventually developed and utilized, the *United States v. Tapia* (2011) decision should be referenced, particularly if treatment and sentencing decisions are being made with the same instrument. The *Tapia* decision ruled that a federal court may not impose or lengthen a prison sentence for the purposes of rehabilitation. As such, the goals, role, and influence of a needs assessment as it relates to sentencing decisions should be made clear early in the development phase of the corresponding instrument.

5. What information should be included in a risk assessment tool?

a. How should recidivism be defined and measured?

Recommendation:

There are a number of ways of defining and measuring offender recidivism. The goals of this project support the use of subsequent arrests as a measure of recidivism. The strengths and weakness of this and other potential measures, such as reconviction or self-report, must be considered.

Justification:

A practical methodology for the purposes of the present policy and evaluation is to measure recidivism as the time until first arrest. The feasibility of measurement, relevance to criminal justice processes, and availability of necessary data makes it a viable option. Furthermore a recent study of recidivism in Virginia found that special attention should be paid to those who recidivate rapidly; an advanced statistical analysis found that the number of crimes one would have committed while un-incarcerated is the inverse

of the length of time until subsequent arrest (Netter, 2007). As such, time until first arrest appears to be an important recidivism measure. However, this approach is not without shortcomings; for example, the first arrest is not necessarily the first offense committed—just the first discovered; additionally it is possible that the police may have arrested the wrong individual.

Many analyses of recidivism have calculated relapse rates by determining the proportion of a cohort that was rearrested within a stated time period after release; however this method generally captures population rates, and is not appropriate for measuring individual recidivism. For example, Durose and colleagues (*BJS Special Report*, 2014) reported that within three years 67.8 percent of released prisoners were rearrested. While this fact is interesting and captures information at the aggregate level, scholars such as Travis (2005) argue that such recidivism measures are inadequate. Instead, Travis advocates for highlighting the frequency and severity of reoffending, stating that "a measure that does not differentiate between crimes committed the day following release and crimes committed two years after release provides a weak foundation for policymaking" (2005:100). As such, it does not seem that quantifying the frequency or proportion of arrest for a given time period would be a useful measure for the present research.

Another potential measure of recidivism is reconviction. This measure would provide specificity regarding the offense, which could be used to update risk (e.g. if the new offense was violent, drug-related, etc.); this would also provide greater certainty of guilt, as the individual would have been convicted, rather than just arrested. However, reconviction measures are likely to miss many earlier offenses for which the individual was either not prosecuted and/or found guilty. Additionally, due to the lengthy criminal justice process, the information about time until first offense would be lost. Finally, the Criminal History records that will most likely be used to measure recidivism are much less complete for court processing of offenders than they are for arrests (booking).

Re-incarceration, another potential measure, might provide a wider net of infractions (ranging from new offenses to technical violations); however this measure also does not necessarily ensure that it is capturing the first (or one of the first) offense(s). Many offenses drop out of the criminal justice system before a person is incarcerated for that offense, leading to an underestimate of recidivism. Re-incarceration might be useful for assessing the effects of sentencing decisions on the correctional system, but it is not as appropriate for assessing effects on reoffending and public safety.

Finally, self-report data might best capture the true first reoffending incident and be most useful in developing risk assessment tools, but this method is labor intensive, and requires compliance and honesty.

b. What information should be used in the tool to predict recidivism?

Recommendation:

Maryland's initial risk assessment instrument should value predictive power over its ability to explain what factors cause recidivism and how to manipulate them. Both acquired and ascriptive characteristics (e.g. age and gender) should also be considered given their ability to inform the instrument's predictions; inclusion of certain characteristics should only come after the legal and ethical appropriateness of including these elements has been discussed by the Commission.

Justification:

Some academics argue in favor of using astructural models (i.e. models that are purely for the purpose of *predicting* phenomena, rather than explaining *why* the phenomena occur). Their argument is that the primary goal of risk assessment is forecasting and not explanation, therefore predictors that improve forecasting accuracy should be included in the model even if the relationship between the predictor and the outcome makes little intuitive sense (Berk & Bleich, 2014).[3] As such, Berk has argued that astructural models are better for short term forecasts, and structural with long term, as the causal process is necessary for any sort of policy-relevant manipulation (Berk, 2008). This reasoning suggests that the development or integration of an astructural model would likely suit the interests of Maryland sentencing reform.

There is a wide breadth of research that describes which variables should be considered for inclusion when developing an astructural model. This body of research provides support for the use of both ascriptive (i.e. static) and achieved (i.e. dynamic) characteristics. Pulling from the experiences of three states that have integrated risk assessment into sentencing practices—Virginia, Pennsylvania, and Missouri—we see that all three utilize static personal factors, offense-specific components, and dynamic individual variables in generating risk assessments.[4]

Virginia's "Worksheet D", which identifies low-risk offenders for diversion utilizes six offender characteristics to determine recidivism risk: offense type, whether the offender is charged with an additional offense, offender characteristics (gender, age, employment, and marital status), whether they were arrested in the past eighteen months, prior felony convictions, and prior adult incarceration (Monahan & Skeem, 2014). Evaluations of Virginia's instrument found that prior record and gender should be weighted more heavily than other components (Kleiman et al., 2007). Like Virginia, Pennsylvania also incorporates ascriptive and achieved characteristics. Pennsylvania's items include gender, age, county, prior arrests, whether the individual is a property offender, and the offense gravity score. Lastly, Missouri has also relied both on ascriptive and achieved factors, including age, education, employment, criminal

[3] Risk assessment instruments may be "structural" or "astructural"; the structural models attempt to represent the causal mechanisms behind recidivism, whereas astructural models utilize associations and their purely predictive (rather than explanatory) value. For example if including a particular variable improves the predictions produced by an instrument, an astructural model would include it, regardless of whether the relationship between the variable and the outcome could be explained (Berk & Cooley, 1987).

[4] N.B. The core issue in the prediction of recidivism is *the probability of an individual's recidivism, given a particular characteristic*; the concern is not the probability of a characteristic, given that an individual recidivated. It is important to be mindful of this difference when considering which variables to include in the model and the potential controversy that might come from their inclusion.

history, and prior escape (Wolff, 2006). Specifically, Missouri's automated risk assessment tool includes offender based variables (e.g. prior criminal history and current offense details) as well as the actual time served by similar offenders; additionally, Missouri incorporates clinical risk though the Sentencing Assessment Report, which contains victim impact, offense history, and other dynamic variables (Hyatt et al., 2011).

Additionally, research from other aspects of the criminal justice process—including pretrial and post-release supervision—can help inform the predictive value of certain factors on offending. A summary report on the current state of pretrial risk assessment identified six of the most common validated pretrial risk factors that have appeared in studies since 2003. These variables included prior failures to appear for court, prior convictions, if the present charge is a felony, being unemployed, history of drug abuse, and having a pending case (Mamalian, 2011). Other research, such as Berk and colleagues' evaluation of an astructural random forests instrument used to predict murderous conduct by offenders on probation and parole over a two-year period in Philadelphia found the most significant predictors to be age of the offender, age of first contact with the adult court system, and number of prior convictions involving a firearm (Berk et al., 2009).

Thus, while a tool developed for sentencing in Maryland will need to be calibrated to reflect the state's unique laws, crime issues, and offender characteristics, it would be most efficient to begin by working with factors already validated in other more experienced states.

When deciding which variables to incorporate into the instrument, one must consider the controversial nature of the use of particular characteristics (or combinations of characteristics) in risk assessment. For example, while the use of racial or ethnic characteristics to determine risk may be seen as unconstitutional or morally questionable, the morality of using other factors such as criminal history or employment and education, is less clear. Some critics have warned that racially disparate policing and enforcement practices cause specific racial groups to accumulate lengthier criminal histories, despite comparable rates of offending for particular crimes (Jannetta et al., 2014). Additionally, Attorney General Eric Holder warns that using "static, historical offender characteristics such as education level, employment history, family circumstances and demographic information…is a dangerous concept that will become much more concerning over time as other far reaching sociological and personal information unrelated to the crimes at issue are incorporated into risk tools" (Holder, 2014: 7).

Conversely, advocates for the use of current risk assessment tools, such as former New Jersey Attorney General Anne Milgram, argue that the use of objective, quantifiable factors provides a more objective, efficient, and fair approach to the administration of justice (Milgram, 2014). Furthermore, Milgram and her colleagues at the Arnold Foundation analyzed over one million criminal cases from across the United States, with the purpose of identifying factors that best predict violence, reoffending, or skipping court. They found that the most controversial social variables (education level, socioeconomic status, and neighborhood) were not among the best predictors (Milgram, 2014). Thus, it does not appear that the current issues being raised regarding the objectivity of risk assessment instruments should discourage practitioners from considering variables beyond those specific to the offense at hand. That said, it is

important for members of the Commission to take the time to discuss and explore the legal and ethical boundaries that might restrict their own instrument development.

The Commission should discuss the possibility of incorporating a racial impact statement to accompany any risk assessment instrument adopted. This statement could preemptively describe potential unintended consequences, such as sentencing disparity, which could result from implementing the proposed instrument. The Commission may choose to conduct a preliminary statistical analysis in conjunction with this statement to demonstrate why and how racially disparate trends might emerge, in addition to the policy's public safety implications.

6. How should predictors be weighted to obtain a risk score?

Recommendation:

Predictors should be weighted through data training processes. By calibrating the instrument with a large sample of historical data, and evaluating and revalidating the tool regularly, it should be clear which variables should be weighted more heavily. Again, this is a situation that can be informed by other states' experiences.

Justification:

The initial instrument (regardless of the statistical approach utilized) should be developed using three basic steps. First a classification scheme should be developed using pretrial predictor data that include potential predictors, and the outcome class of interest. Next the forecasting accuracy of the model should be evaluated using test data from the same jurisdiction that include the same predictors and outcome classes. If the test data's forecasting ability is satisfactory, then the model may be used in situations when the predictors are known, but the outcome is not (Berk & Bleich, 2014). Finally, instruments should be re-validated over time to ensure that particular variables do not need to be re-weighted or otherwise adjusted as the criminal or social landscape of the jurisdiction changes, and to ensure that the instrument is correctly being implemented by well-trained staff (Andrews et al., 2006; Lowenkamp et al., 2004).

Identification of the initial weights can occur when the model is being assessed using historical data. If an "off the shelf" instrument is being utilized in an astructural capacity, then weights can be readjusted until maximum predictive ability has been achieved. If an individualized instrument is being developed from the ground up, then development will rely more heavily on other states' past experience and the existing literature.

Depending on the type of model developed, weighting of variables may require additional precautions. For example, classification and regression tree (CART) modeling involves constructing profiles of individuals associated with different outcome classes (e.g. high risk and low risk). These classifications often need to be adjusted when new data are added into the model; as such, the weights must consistently be inspected and updated (Berk & Bleich, 2014). This issue can be dealt with if a "random forests" approach to CART is utilized; these models are more stable, and more adaptive to new data, but require a more advanced internal IT structure (Berk & Bleich, 2014).

7. How should a risk assessment tool be implemented?

a. Should a new instrument be developed or can an existing tool be modified?

Recommendation:

This decision should be guided by the time and budget restrictions placed on the project. Developing an instrument *de novo* is more likely to meet the specific needs and goals of the Commission, as well as be tailored to Maryland's unique offender population. However this approach is likely to be more expensive and time consuming. The Commission may instead consider working with a team of professionals to identify an "off the shelf" instrument or a tool developed and implemented for another jurisdiction, and recalibrate it to fit Maryland's needs and population. Maryland's Department of Public Safety and Correctional Services has indicated that it is in the process of transitioning to the use of LSI-R for placement based decisions, in lieu of the current risk assessment tool.[5] As this transition is currently taking place, at the present time Maryland is not in a position to adapt a risk assessment instrument internally from another agency.

Justification:

In the past, some jurisdictions have attempted to implement "off the shelf" models, or transplant instruments developed in other states. This has proven problematic in some cases, as the model was validated on a different population, without considering differences between the target population's characteristics (Bechtel et al., 2011). Jurisdictions should instead work to select instruments that fit their assessment needs and ensure that they have been properly validated for use with their unique population (Casey et al., 2011). A 2010 national survey by the Vera Institute of Justice reported that of forty-one responding agencies, twenty reported using a state-specific tool; this indicates that the instrument was either developed specifically for that state, or the state modified a generic tool such as the LSI-R.

When deciding whether to develop a new tool or modify an existing instrument (such as the LSI-R) factors including the time, money, and skills required to develop and validate a new instrument can be prohibitive; as such, many opt to implement a pre-existing model. Additionally, another agency within the jurisdiction may already have an actuarial instrument in place (e.g. probation, parole); this instrument may be modified to assess sanction-related risk, and allow for continuity and communication across the jurisdiction's criminal justice system (see Latessa & Lovins, 2010). However, it must be noted that different variables may need to be added, dropped, or re-weighted depending on the criminal justice process they are being used to predict (e.g. pre-trial flight compared to post-release offending) (Bechtel et al., 2011).

Maryland currently has risk assessment instruments used at other stages of the criminal justice process, specifically for probation, prison entry, and parole. Maryland's Department of Public Safety and

[5] http://dbm.maryland.gov/agencies/operbudget/FY2015Testimony/Q00C01.pdf

Correctional Services has indicated that it is in the process of transitioning to using LSI-R for placement based decisions, and is therefore not currently in a position to adapt a risk assessment instrument internally from another agency. The LSI-R in particular includes many dynamic factors (i.e. needs), which makes it more appropriate in a corrections and programming capacity than for making decisions on whether to incarcerate someone. Additionally, during the May 2011 Guidelines Subcommittee meeting the option of modifying an existing risk assessment instrument to use at sentencing was discussed; Subcommittee member Richard A. Finci noted that what is being predicted by the risk assessment instrument may impact measures in subsequent instruments (i.e. parole) and that the different tools predict different outcomes.

Ultimately the goals for the initial risk based sentencing instrument are to effectively identify low-risk offenders for diversion from incarceration sentences. If time and budget allow, it would be ideal to utilize available information (e.g. from the Sentencing Commission, shared records, etc.) to develop a sophisticated, well-validated tool, such as a CART model. In this situation, the model would serve the purpose of identifying low-risk offenders, who could be selected for diversion programs.

b. Are there any types of offenders that are not to be included in the instrument?

Recommendation:

Consistent with Recommendation Three, the initial instrument should specifically target low-risk nonviolent offenders for diversion. Within this group, all offenses eligible for diversion from imprisonment (i.e. probation-eligible) should be considered in the development of the instrument. Through data review, instrument calibration, and sensitivity analysis, unique pools of offenders that may be a poor fit for risk assessment will be revealed.

Justification:

In their own analysis, the Pennsylvania Commission on Sentencing opted to study offenders convicted of driving under the influence (DUI) separately from other offenders, stating that this is because "they differ from non-DUI offenders on several dimensions" (PA Commission on Sentencing, *Interim Report*). On the other hand, Virginia developed two separate instruments, with one being a separate sex offender tool. The primary risk assessment instrument is used to identify appropriate nonviolent, low-risk offenders for diversion from prison; the secondary tool identifies high-risk sex offenders for longer periods of incapacitation through incarceration (Kern & Owens, 2004). Unlike the primary tool, the sex offender instrument works to identify *high-risk* individuals and adjust guidelines upward. To develop the instrument, a separate study was conducted to identify, weight, and validate characteristics common among sex offenders who recidivate (Soulé & Najaka, 2013).

If Maryland's instrument is to be used to determine who on the margins for incarceration should be given a diversionary sentence, then violent sex offenders are of no concern, as they are unlikely to be eligible for a non-incarceration sentence. Additionally, it is clear from other jurisdictions' experiences with implementing "off the shelf" tools that populations are not uniform. As such, rather than excluding

particular types of offenders because it was done in other states, during the instrument development/calibration and validation, developers should be wary of offender groups that are atypical, or deviate from trends depicted by others. If DUI or other particular groups of offenders appear in the data to be different from the norm, and unable to be classified in a general instrument, then they should be excluded. Furthermore, some groups may be excluded for the sake of "justice". For example, while some violent offenders may recidivate at lower rates than nonviolent offenders, they are unlikely to be considered for diversionary sentences for punitive reasons.

c. Should the tool be voluntary or compulsory?

Recommendation:

Consistent with the Maryland Sentencing Guidelines, the tool should serve in an advisory capacity.

Justification:

As stated in recommendation number five, clinical judgment and judicial discretion may play an important role in the decision regarding how best to use an individual's risk score. This would allow judges to consider situational and contextual factors that may not be a part of the instrument. Additionally, given that the state's sentencing guidelines are not presumptive, and instead function in an advisory capacity, it would be logical for the risk assessment tool guiding sentencing decisions to do the same (particularly as the threshold for incarceration/diversion decisions will likely be influenced by Sentencing Guidelines standards).

d. Should the risk assessment instrument be integrated with the existing Maryland Sentencing Guidelines or function as a stand-alone tool?

Recommendation:

The instrument should be developed in such a way that it is consistent with the Maryland Sentencing Guidelines. As the Guidelines are voluntary/advisory, the risk assessment instrument should be able to operate independently, but the recommendations should be consistent with the Guidelines regarding which cases are suitable for diversionary sentences.

Justification:

Virginia was the first state to explicitly incorporate risk assessment into its sentencing guidelines. In Virginia, an incarceration-bound offender receiving a score below a particular threshold may be recommended for an alternative sanction. The instrument is only scored for offenders who meet the sentencing guidelines recommendation for incarceration (i.e. probation cases are not considered for diversion), and a criminal history of only nonviolent offenses.

It appears critical that the risk assessment threshold and sentencing guidelines be in agreement regarding where deferrable offenders fall; the instruments should be calibrated in such a way that the risk score

threshold aligns with offenders who are likely to receive incarceration sentences, but eligible for probation.

References

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

Austin, J.F. 2004. The proper and improper use of risk assessment in corrections. *Federal Sentencing Reporter*, 16(3) 1-6.

Bechtel, K., Lowenkamp, C., and Holsinger, A. 2011. Identifying the predictors of pretrial failure: A meta-analysis. Final Report

Berk, Richard A. 2008. Forecasting methods in crime and justice. *Annual Review of Law and Social Science (Annual Reviews Press)* 4, 219–238.

Berk, R., and Bleich, J. 2014. Forecasts of violence to inform sentencing decisions. *Journal of Quantitative Criminology*, *30*(1), 79-96.

Berk, R. A., and Cooley, T. F. 1987. Errors in forecasting social phenomena. *Forecasting In The Social And Natural Sciences* (pp. 247-265). Springer: Netherlands.

Berk, R.A., Sherman, L., Barnes, G., Kurtz, E., and Ahlman, L. 2009. Forecasting murder within a population of probationers and parolees: a high stakes application of statistical learning. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 172(1), 191-211.*

Bushway, S., and Smith, J. 2007. Sentencing using statistical treatment rules: what we don't know can hurt us. *Journal of Quantitative Criminology*, *23*(4), 377-387.

Casey, P. M., Elek, J. K., Warren, R. K., Cheesman, F., Kleiman, M., and Ostrom, B. 2014. Offender risk and needs assessment instruments: A primer for courts. Williamsburg, VA: National Center for State Courts. Retrieved from http://www.ncsc.org/~/media/Microsites/Files/CSI/BJA%20RNA%20Final%20Report_Combined%20Files%208-22-14.ashx.

Cooke, D.J. and Michie, C. 2010. Limitations of diagnostic precision and predictive utility in the individual case: A challenge for forensic practice. *Law and Human Behavior, 34* 259-74.

Cullen, F.T. and Gendreau, P. 2000. Assessing correctional rehabilitation: Policy, practice, and prospects. *Policies, Processes, and Decisions of the Criminal Justice System*, 3, *109-175.*

Durose, M. R., Cooper, A. D., and Snyder, H. N. 2014. Recidivism of Prisoners Released in 30 States in 2005: Patterns from 2005 to 2010. *Bureau of Justice Statistics*.

Gottfredson S. and Gottfredson, D. 1986. Accuracy of prediction models. In: Blumstein A, Cohen J, Roth J, Visher C (eds) *Criminal Careers and ''Career Criminals''*. National Academy Press, Washington, DC, pp 212–290.

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

Hart, S.D., Michie, C., and Cooke, D.J. 2007. Precision of actuarial risk assessment instruments: Evaluating the 'margins of error' of group v. individual predictions of violence. *British Journal of Psychiatry, 190* 60-5.

Heilbrun, K. 2009. *Evaluation for Risk of Violence in Adults.* New York: Oxford University Press.

Hyatt, J.M., Chanenson, S.L. and Bergstrom, M.H. 2011. Reform in motion: The promise and perils of incorporating risk assessments and cost-benefit analysis into Pennsylvania sentencing. *Villanova Public Law and Legal Theory Working Paper Series.*

Jannetta, J., Breaux, J., Ho, H., and Porter, J. 2014. *Examining Racial and Ethnic Disparities in Probation Revocation: Summary and Findings from a Multisite Study*. Washington, DC: Urban Institute.

Kern, R.P. and Farrar-Owens, M. 2004. Sentencing guidelines with integrated offender risk assessment. *Federal Sentencing Reporter, 16(3)* 165-9.

Kleiman, M., Ostrom, B.J., and Cheesman, F.L. 2007. Using risk assessment to inform sentencing decisions for nonviolent offenders in Virginia. *Crime & Delinquency, 53,* 106-32.

Latessa, Edward J. (2004). The challenge of change: Correctional programs and evidence-based practices. *Criminology & Public Policy, 3(4)* 547-60.

Latessa, E.J. and Lovins, B. 2010. The role of offender risk assessment: A policy maker guide. *Victims & Offenders, 5(3)* 203-19.

Lownkamp, C.T. and Latessa, E.J. 2004. Understanding the risk principle: How and why correctional interventions can harm low-risk offenders. [Technical report]. *Topics in community corrections, (pp. 3-8).* Washington, DC: U.S. Department of Justice, National Institute of Corrections.

Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. 2004. Empirical evidence on the importance of training and experience in using the Level of Service Inventory-Revised. *Topics in Community Corrections,* 49-53.

Mamalian, C.A. 2011. State of the science of pretrial risk assessment. Pretrial Justice Institute. Retrieved from: http://www.pretrial.org/download/risk-assessment/PJI%20State%20of%20the%20Science%20Pretrial%20Risk%20Assessment%20(2011).pdf

Milgram, A. 2014. Why smart statistics are the key to fighting crime. *TED Talk transcript.* Retrieved from: https://www.ted.com/talks/anne_milgram_why_smart_statistics_are_the_key_to_fighting_crime/transcript?language=en

Monahan, J., & Skeem, J. L. (2014). Risk Redux: The Resurgence of Risk Assessment in Criminal Sanctioning. *Federal Sentencing Reporter*, *26*(3), 158-166.

Netter, B. 2007. Using group statistics to sentence individual criminals: an ethical and statistical critique of the Virginia Risk Assessment Program. *The Journal of Criminal Law and Criminology*, 699-729.

Pew Center on the States. 2011. Risk/Needs Assessment 101: Science reveals new tools to manage offenders. *Public Safety Performance Project Issue Brief.* Retrieved From: http://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2011/09/20/riskneeds-assessment-101-science-reveals-new-tools-to-manage-offenders

Skeem, J.L. and Monahan, J. 2011. Current directions in violence risk assessment. *Current Directions in Psychological Science*; Virginia Public Law and Legal Theory Research Paper No. 2011-13.

Soule, D.A. and Najaka, S.S. 2013. Paving a path to informed sentencing decisions. *Federal Sentencing Reporter, 25(3)* 181–184.

Starr, S.B. (2014, August 11). Sentencing by the Numbers. *The New York Times*, p. A17.

Travis, J. 2005. *But They All Come Back: Facing the Challenges of Prisoner Reentry.* The Urban Institute.

VanNostrand, M. and Lowenkamp, C.T. 2013. Assessing pretrial risk without a defendant interview. *Laura and John Arnold Foundation*. Retrieved from: http://www.arnoldfoundation.org/sites/default/files/pdf/LJAF_Report_no-interview_FNL.pdf

Visher, C. A. 1987. Juror decision making: The importance of evidence. *Law and Human Behavior, 11*, 1-17.

Wolff, J. 2006. Risk, fear, blame, shame and the regulation of public safety. *Economics and Philosophy, 22*: 409-27.