Achieving Justice: Reducing Racial Bias in Criminal Justice Risk Assessments

Risk-Need Assessments (RNAs) have become crucial decision-making tools across the criminal justice system. Built to reduce bias and increase accuracy in release and programming decisions, these tools provide a standardized assessment of individuals’ risk of reoffending, failure to appear, or infraction behavior, as well as identify needs to be targeted by programming (Andrews & Bonta, 2010). Risk levels (i.e., Low, Moderate, & High-Risk) are then used to determine intensity of supervision, with higher risk levels corresponding to more restrictive/intense supervision. Accurately identifying individuals’ risk levels is crucial to maintaining public safety and providing rehabilitative programming effectively. For example, among federally incarcerated individuals, study findings showed that 49.3% of high-risk persons recidivated within one year of release compared to just 4% of low-risk individuals (Cohen et al., 2016). Similarly, other studies show that providing intensive treatment that is typically reserved for high-risk individuals to low-risk ones increases recidivism (Lowenkamp & Latessa, 2004). Despite the clear need for accurate assessments, concerns have been raised regarding their accuracy across racial/ethnic groups. Specifically, recent research has found that racial/ethnic minorities are more likely to be overclassified as High-Risk, creating an inaccurate and biased prediction (Vincent & Viljoen, 2020).

**Summary**

Risk-Need Assessments (RNAs) are important tools used in the criminal justice system to make decisions about release and programming. The assessments are meant to help overcome the biases and inaccuracies inherent in human decision making. However, when not carefully accounted for, tools’ risk prediction can extend forward racial/ethnic biases that persist in other parts of the criminal system. Subsequently, some policy and advocacy groups have called for their removal. This would not solve the issue of racial/ethnic bias as RNAs merely reflect the system and community factors that the sources are disparate treatment are rooted in.

This brief addresses this issue of ‘Bias in Bias Out’ by providing and overview of methodological advancements in RNA development. Specifically, we speak to the importance of gathering local data to identify dynamic assessment items that are apt to reflect individuals’ progression through the criminal justice system rather than factors inextricably linked to demographics.

We find that using local data helps to create tools that demonstrate greater accuracy by capturing variations in jurisdiction’s laws, policies, and populations. Further, we find that dynamic items help to reduce racial/ethnic bias while providing a better understanding of human behavior. Using these methods in tandem maximizes accuracy while and does not extend forward the biases in other parts of the community and criminal justice system.

When assessment's results are applied to policy, bias and overclassification lead to imposing overly restrictive supervision, unnecessary programming, and have the potential to worsen disparities through greater system contact, detention, and monitoring (Loughran et al., 2013).

Some have argued that bias is ‘baked in’ to assessment results, as tools are developed on predominantly White, male populations (Mayson, 2018, Van Eijk, 2017). Further, these tools are commonly created in one location and applied ‘off-

the-shelf', meaning they are deployed without adjusting for local population distinctions. While still a common practice today, overclassification is more likely in ‘off-the-shelf’ tools, and this issue is more severe when adopted by jurisdiction that is demographically diverse (Picard-Fritsche et al., 2017).

Decades of research have determined the evidence-base of risk-need assessments and their improved accuracy compared to human practitioners and judicial decision makers (Lin et al., 2020). Unfortunately, some researchers and advocacy groups have suggested the removal of assessments from justice practice (Green, 2020; Schwerzmann, 2021), returning to the idiosyncratic discretion of subject matter experts. Fortunately, efforts to combat bias have expanded.

Localization

Localization is the first step in combatting bias. Sometimes termed ‘homegrown’, RNAs developed using data gathered from the specific, local jurisdiction are found to be more accurate for the population (Duwe, 2024). One rationale for the improved accuracy is that each jurisdiction has unique laws and enforcement patterns that may perpetuate implicit biases differently across jurisdictions. This variation can reduce accuracy when an off-the-shelf tool is applied in a new location and is reflected in tools’ rates of overclassification.

To illustrate, we provide findings from a recent evaluation of the Ohio Risk Assessment System (ORAS; Latessa et al., 2010) used with a community corrections sample in a Midwest State. Created in Ohio and implemented off-the-shelf, the ORAS was identified to have a ‘Weak’ predictive accuracy rating.[[1]](#footnote-2) Figure 1 provides a scatter plot, with ORAS risk scores along the horizontal and recidivism probabilities on the vertical axis. Each line provides the recidivism rate associated with White and Black individuals. As demonstrated, the White trend line rises at a greater rate, increasing its separation from the Black trend line. This plot not only demonstrates that the ORAS prediction is superior for White individuals, but the gap between plot lines also indicates that, despite receiving the same risk score, higher risk Black individuals recidivate at an 8-10% lower rate than their White counterparts. This is a clear example of overclassification and, as described, is more likely to result when a tool is developed in one location and applied in a new jurisdiction with a substantially distinct population.

**Figure 1: ORAS Overclassification (N=120,212)**



Dynamic Items

Another source of bias is found in the type of items used. Many contemporary tools rely heavily on criminal history measures (Picard-Fritsche et al., 2017). Criminal history items are ‘static’, meaning that they cannot change over time. When more of these items are used in an assessment, the greater the difficulty for an individual to lower their risk score or demonstrate reduced risk over time. Unlike criminal history items, dynamic items allow individuals to reduce their risk over time by demonstrating behaviors and thought patterns conducive to successful community reintegration (Andrews & Bonta, 2010). Hence, tools that capture a greater number of dynamic indicators (i.e., changeable factors) provide an assessment of needs, or risk factors that can be reduced through programming and service provisions.

In contrast, tools that rely heavily on criminal history and other static items, have a greater tendency to overestimate risk for racial/ethnic minorities (Miller et al., 2022). This tendency is associated with law enforcement and prosecutorial practices that produce Disproportionate Minority Contact (DMC) (Butler et al., 2022). DMC refers to greater rates of contact with the justice system among individuals of a specific minority group. Notably, these groups are more likely to reside in high crime areas with greater law enforcement patrols. Therefore, heavy reliance on criminal history items not only reflects these practices in risk scores but can lead to minority individuals’ placement in higher levels of supervision. More stringent supervision only serves to perpetuate DMC, as contacts with the criminal justice system for high-risk individuals are more frequent and more restrictive (Andrews & Bonta, 2010; Eckhouse et al., 2019).

However, dynamic needs indicators (e.g., family conflict, substance abuse, & mental health issues) are not impacted by DMC. Therefore, tools that make use of a greater number of dynamic items reduce an assessment’s reliance on, and bias potency of, criminal history items in predicting risk (Miller et al., 2022). When created with local data, tools with greater dynamic content are found to reduce bias while increasing tool accuracy (Hamilton et al., 2022). In a recent example, the Personal Recognizance Interview & Needs Screen (PRINS) was developed for King County Washington in 2017. In an effort to improve upon the mostly static, criminal history-based pretrial tools on the market, King County sought to collect dynamic indicators post-booking, to create a more accurate assessment of pretrial risk. Figure 2 provides a scatter plot indicating the positive impacts (PRINS), where White and Non-White defendants demonstrate near equal rates of offending (vertical axis) across the tool’s risk scores (horizontal axis). In contrast to Figure 1, the PRINS prediction identifies an absence of overclassification among racial/ethnic minorities.

**Figure 2: PRINS Prediction by Race/Ethnicity (N=28,147)**



Notably, when compared to the three major pretrial assessment tools, which are composed mostly of criminal history indicators, the PRINS identified greater predictive accuracy and near equality of prediction across race/ethnicity groups (Hamilton et al., 2024). While representing only one example, Figure 2 illustrates the combined impact that dynamic item inclusion and localization can have when designing and updating risk-needs assessments.

Conclusion

In conclusion, racial/ethnic bias in risk-need assessment can often be viewed as a reflection of local laws, policing, and prosecutorial practices. While not the cause, *an assessment provides a mirror that reflects bias inherent in the law enforcement and prosecutorial practices found within a jurisdiction*. Deemed either intentional or unintentional by practitioners, advocates, and decision-makers, these biases are rooted in the community and said biases are merely measured, or observed, by the risk-needs assessment. These biases are found in progressive cities, moderate suburbs, and even conservative towns.

While some have called for the removal of risk-need tools as a solution to observed rates of bias found in assessment results (Green, 2020; Schwerzmann, 2021), this is not an effective solution to reduce bias within the criminal justice system. As indicated, an assessment tool is made up of measures gathered from the community, and is not in itself biased, but reflective of the bias found in that jurisdiction’s justice system. Removing the tool turns a blind eye to this issue and prevents the observation of bias within agency decision making. Further, removing assessment tools returns an agency to idiosyncratic decision making that has been routinely found to reduce decision making accuracy and produce greater, not lesser, rates of bias (Ægisdóttir et al., 2006; Cohen et al., 2016; Grove et al., 2000; Skeem & Monahan, 2011).

In response to these calls, recent research has uncovered methods of reducing bias using local data and a greater inclusion of dynamic items in the development of risk-need tools. As described here, localization ensures that the items and scores are reflective of the jurisdiction’s population. By accounting for variations found in the local policies and laws, homegrown tools have demonstrated greater accuracy in prediction and reflect a population’s diversity. Further by reducing the impact of criminal history indicators, a greater inclusion of dynamic items provides a better understanding of the individual and their behavior and has the positive effect of reducing bias. When not accounted for, these inherent biases within criminal history measures increase the risk of overclassification and lead to greater rates of DMC within the criminal justice system. Further, when identifying dynamic indicators that are uniquely predictive of the local population, greater accuracy is achieved.

References

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., ... & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. The counseling psychologist, 34(3), 341-382.

Andrews, D. A., & Bonta, J. (2010). The psychology of criminal conduct. Routledge.

Butler, L. C., Hamilton, Z., Krushas, A. E., Kigerl, A., & Kowalski, M. (2022). Racial bias and amelioration strategies for juvenile risk assessment. Handbook on inequalities in sentencing and corrections among marginalized populations, 70-118.

Cohen, T. H., Pendergast, B., & VanBenschoten, S. W. (2016). Examining overrides of risk classifications for offenders on federal supervision. Fed. Probation, 80, 12.

Cohen, T. H., Lowenkamp, C. T., & VanBenschoten, S. W. (2016). Does change in risk matter? Examining whether changes in offender risk characteristics influence recidivism outcomes. Criminology & Public Policy, 15(2), 263-296.

Duwe, G. (2024). Evaluating bias, shrinkage and the home-field advantage: Results from a revalidation of the MnSTARR 2.0. Corrections, 9(1), 20-42.

Eckhouse, L., Lum, K., Conti-Cook, C., & Ciccolini, J. (2019). Layers of bias: A unified approach for understanding problems with risk assessment. Criminal Justice and Behavior, 46(2), 185-209.

Green, B. (2020, January). The false promise of risk assessments: epistemic reform and the limits of fairness. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 594-606).

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.

Hamilton, Z., Kowalski, M. & Kigerl, A. (2022) Prediction is Local: The Benefits of Optimization. Justice Quarterly, 39:4, 722-744.

Hamilton, Z., Ursino, J., Kigerl, A. (2024). Personal Recognizance Interview and Needs Screen (PRINS): Development of a Pretrial Risk Assessment. Journal of Criminal Justice.

Latessa, E. J., Lemke, R., Makarios, M., & Smith, P. (2010). The creation and validation of the Ohio Risk Assessment System (ORAS). Fed. Probation, 74, 16.

Lin, Z. J., Jung, J., Goel, S., & Skeem, J. (2020). The limits of human predictions of recidivism. Science advances, 6(7), eaaz0652.

Loughran, T. A., Mulvey, E. P., Schubert, C. A., Fagan, J., Piquero, A. R., & Losoya, S. H. (2009). Estimating a dose‐response relationship between length of stay and future recidivism in serious juvenile offenders. Criminology, 47(3), 699-740.

Lowenkamp, C. T., & Latessa, E. J. (2004). Understanding the risk principle: How and why correctional interventions can harm low-risk offenders. Topics in community corrections, 2004, 3-8.

Mayson, S. G. (2018). Bias in, bias out. YAle lJ, 128, 2218.

Miller, W. T., Campbell, C. A., Papp, J., & Ruhland, E. (2022). The contribution of static and dynamic factors to recidivism prediction for Black and White youth offenders. International journal of offender therapy and comparative criminology, 66(16), 1779-1795.

Picard-Fritsche, S., Rempel, M., Tallon, J. A., Adler, J., & Reyes, N. (2017). Demystifying risk assessment. *Center for Court Innovation.*

Schwerzmann, K. (2021). Abolish! Against the use of risk assessment algorithms at sentencing in the US criminal justice system. Philosophy & Technology, 34(4), 1883-1904.

Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. Current directions in psychological science, 20(1), 38-42.

Van Eijk, G. (2017). Socioeconomic marginality in sentencing: The built-in bias in risk assessment tools and the reproduction of social inequality. Punishment & Society, 19(4), 463-481.

Vincent, G. M., & Viljoen, J. L. (2020). Racist algorithms or systemic problems? Risk assessments and racial disparities. Criminal justice and behavior, 47(12), 1576-1584.

1. Predictive accuracy was measured vias the Area Under the Curve (AUC) statistic, and established ranges indicate 0.5-0.55 as ‘Negligible’, 0.55-0.63 as ‘Weak’, 0.64-0.70 as ‘Moderate’, and 0.71 and above as ‘Strong’. This evaluation of the ORAS reported AUCs of 0.63. [↑](#footnote-ref-2)