



THE OHIO STATE UNIVERSITY

PROGRAM ON DATA AND GOVERNANCE

A program of the Moritz College of Law and the
Translational Data Analytics Institute

Aligning Algorithmic Risk Assessments With Criminal Justice Values

Dennis D. Hirsch, Professor, Moritz College of Law (Principal Investigator)

Jared Ott, Senior Research Associate (former)

Angie Westover-Munoz, Program Manager

Chris Yaluma, Senior Research Associate



Aligning Algorithmic Risk Assessments With Criminal Justice Values

**The Ohio State University
Program on Data and Governance
A Program of The Moritz College of Law and
the Translational Data Analytics Institute**

Dennis D. Hirsch, Professor, Moritz College of Law (Principal Investigator)
Jared Ott, Senior Research Associate (former)
Angie Westover-Munoz, Program Manager
Chris Yaluma, Senior Research Associate

Contents

| | |
|---|-----------|
| 1. Introduction | 5 |
| 2. Methodology | 8 |
| 2.1 Study Design | 8 |
| 2.2 Respondent Demographics | 9 |
| 3. Ohio use of ARAs | 10 |
| 3.1 Ohio’s adoption of risk assessments | 10 |
| 3.2 Ohio Risk Assessment Process | 12 |
| 3.21 Items on PAT and CST | 13 |
| 3.22 Risk Assessments in the Ohio Courts of Common Pleas | 15 |
| 3.23 Variation in use of ARAs across and within counties | 18 |
| 3.3 Judicial views and attitudes towards ARAs tools | 21 |
| 3.31 Overall importance of ARAs for decision-making | 21 |
| 3.32 General views towards risk assessment tools | 24 |
| 3.33 Views on accuracy | 29 |
| 3.34 Judicial views on bias | 31 |
| 3.35 Variation in views across counties | 35 |
| 4. Best Practice, Ohio Practice, and Recommendations | 39 |
| 4.1 Interpreting and Communicating Risk Assessment Tool Results | 39 |
| 4.11 Best practice and Ohio practice | 39 |
| 4.12 Recommendations: Communicating Results | 43 |
| 4.2 Training in the Use of ARAs | 43 |
| 4.21 Best practice and Ohio practice | 44 |
| 4.22 Recommendations: Training of Judges and Assessors | 50 |
| 4.3 Ensuring Due Process | 51 |
| 4.31 Best practice and Ohio practice | 51 |
| 4.32 Recommendations: Aligning ARAs with Due Process Values | 55 |
| 4.4 Oversight | 55 |
| 4.41 Best practice and Ohio practice | 55 |
| 4.42 Recommendations: Oversight and Accountability | 58 |
| 5. Recommendations | 59 |
| 5.1 Communicating Results | 59 |
| 5.2 Training of Assessors and Judges | 59 |
| 5.21 Training of assessors | 59 |

| | | |
|------|---|----|
| 5.22 | Training of judges | 60 |
| 5.3 | Aligning ARAs with due process values | 61 |
| 5.4 | Oversight and accountability | 61 |

6. Conclusion **62**

List of Tables

| | | |
|---|--|----|
| 1 | Demographics of judicial survey respondents | 10 |
| 2 | Accuracy of judges’ responses to how categorical risk scores equate to percent risk of recidivism, according to initial ORAS tool validation | 42 |
| 3 | Who receive ARA, when, and what information do they receive | 53 |

List of Figures

| | | |
|----|--|----|
| 1 | Are you currently provided with ARAs for bail and sentencing decisions? | 15 |
| 2 | Respondents use of tools | 16 |
| 3 | How often do you receive a risk assessment for bail and sentencing? | 17 |
| 4 | Extent judges agree with the statement: “ARAs frequently inform my decision-making.” | 19 |
| 5 | Variation in risk assessment usage across counties | 20 |
| 6 | How important do you consider risk assessments to your bail and sentencing decisions? | 21 |
| 7 | Relevance of ARAs for specific decisions | 22 |
| 8 | Factors that influence weight judges give to risk scores | 23 |
| 9 | Judges’ views towards ARA tools | 25 |
| 10 | Judges believe the current cash bail system is: | 27 |
| 11 | Judges believe the incarceration rates across the judicial system are: | 28 |
| 12 | Judges believe the treatment resources in their jurisdiction are adequate. | 29 |
| 13 | Judges’ views on the accuracy of the tools | 29 |
| 14 | Judges’ views of ARA bias compared to human decision-makers | 32 |
| 15 | Judges’ responses to the question, ”How do you feel ARA tools compared to your own ability to assess defendants’ level of risk?” | 32 |
| 16 | ARAs frequently inform my decision-making by county size | 37 |
| 17 | I trust ARA tools by urban and rural counties | 38 |
| 18 | ARAs allow me freedom to make unpopular decision by urban and rural counties | 38 |
| 19 | Amount of training judges reported receiving regarding ARA tools | 46 |
| 20 | Amount of future training judges reported being interested in receiving | 47 |

| | | |
|----|---|----|
| 21 | Extent to which judges agree with the statement: “I feel well-equipped to understand what ARA scores communicate about defendants.” | 48 |
| 22 | Who provided the training | 49 |
| 23 | The extent to which judges agree with the statement: “ARA tools pose due process concerns.” | 52 |
| 24 | Can ARA scores be challenged? | 54 |

Aligning Algorithmic Risk Assessments With Criminal Justice Values

Dennis D. Hirsch
Professor, Moritz College of Law
The Ohio State University
(Principal Investigator)
hirsch.151@osu.edu

Angie Westover-Munoz
Program Manager
Moritz College of Law
The Ohio State University
westover-munoz.1@osu.edu

Christopher B. Yaluma
Senior Research Associate
Moritz College of Law
The Ohio State University
yaluma.1@osu.edu

Jared Ott
Senior Research Associate
(former)
Moritz College of Law
The Ohio State University

October 17, 2024

Abstract

The use of risk assessment (RA) tools has become a key component of the criminal justice system in the United States. Much of the existing scholarship concentrates on normative and technical aspects of RAs, or on recommendations for their improvement. However, there has been little empirical work on how courts and other criminal justice actors perceive and utilize these tools on the ground. In this study, we provide an in-depth picture of how the Courts of Common Pleas think about and use algorithmic risk assessments. Primarily, we focus on the use of risk assessment tools in Ohio Courts of Common Pleas and compare Ohio practices with best practices highlighted in the literature. To investigate, we surveyed Ohio Courts of Common Pleas judges, probation officers, and court administrators regarding their views on and use of algorithmic risk assessment tools. We further conducted interviews with judges and a diverse array of stakeholders that included victim’s rights, civil liberties, and civil rights groups, as well as public defenders and county prosecutors. The findings show that judges largely see risk assessment tools as essential to their decision-making, with most trusting the tools to improve risk-related judgments. While judges believe these tools are no more biased than humans, about 60% still consider their own judgment superior, even though they acknowledge the tools are generally less biased than human decision-makers. Our findings on Ohio’s use of risk assessment tools are mixed. Judges agree the tools should guide, not dictate, decisions, aligning with best practices. However, many lack sufficient training—a crucial recommendation. We conclude with broad recommendations for enhancing the use of risk assessment tools in Ohio’s judicial system.

1. Introduction

Risk assessment is a “process of using risk factors to estimate the likelihood (i.e., probability) of an outcome occurring in a population” (Garrett and Monahan, 2020, p. 448-449). In the criminal justice context, risk assessment uses statistical or machine learning methods to identify those traits that best predict future arrest, conviction, reincarceration, probation revocation, or other negative criminal justice outcomes (Stevenson, 2018). Once they have identified these traits, risk assessment systems apply them algorithmically to a given individual to produce a risk estimate (Berk, 2021; Stevenson, 2018),¹ leading some to refer to them as “algorithmic risk assessments” (Doleac and Stevenson, 2020). In recent years, as criminal justice priorities have increasingly shifted from retribution to rehabilitation and risk management, a growing number of criminal justice systems have come to use algorithmic risk assessment tools (Chanenson and Hyatt, 2016; Garrett and Monahan, 2020; Stevenson, 2018). Today, courts use risk assessments for setting bail and for post-conviction sentencing; corrections institutions use them to assign security levels during prison intake; and parole boards use them to decide on parole applications and determine the appropriate level of parolee supervision (Berk, 2021; Stevenson, 2018).

The criminal justice system’s soaring use of algorithmic risk assessments has been the subject

¹Megan Stevenson (2018) offers a helpful and clear summary:

Most risk assessment tools currently in use are fairly simple checklist-style tools. These tools take a set of inputs, usually between seven and fifteen, and assign a certain number of points to each input. The points assigned to each input are determined through statistical analyses that evaluate how well each input predicts the outcome. The inputs to a risk assessment algorithm almost always include criminal history or criminal justice-related misconduct. Some also include socio-economic factors such as education level, marital status, or home neighborhood. Age and gender are often included, but race is not. The risk score is then calculated by summing the points assigned to each input. Usually, the risk score is then aggregated to a small group of risk classifications: people with the lowest scores are labeled low risk, those with medium scores are labeled moderate risk, and those with the highest scores are labeled high risk. The decision about what fraction of defendants belong in each bin is a normative one.

In addition to the checklist-style risk assessments described above, there are more complicated methods of evaluating risk that are developed through a method called machine learning. Machine-learned risk assessment tools are designed by a computer itself with a little guidance from the person that develops them. The researcher tells the computer which inputs to use, which outcomes to predict, and which learning method to use. The computer does the rest. Machine-learned risk assessments tend to be black-box mechanisms; it is hard to understand why they yield the predictions that they do. This is because the relationship between the inputs and the risk score is non-linear and varied.

of much controversy. Proponents point to proposed benefits such as improved accuracy in assessing risk (Garrett and Monahan, 2020) and reduced need for pre-trial detention or carceral sentences for those deemed low risk (Stevenson, 2018). Critics maintain that algorithmic risk assessments perpetuate and amplify historic racial biases that have long existed in the criminal legal system (Hill, 2021; Mayson, 2018), and that the scoring systems' opacity prevents defendants from adequately examining or challenging their risk scores and so raises important Due Process concerns (Stevenson, 2018). Civil rights organizations and others that once supported the use of risk assessments as a means of reducing mass incarceration have increasingly withdrawn their support or called for changes to the way that these systems are designed and implemented.

Given the controversy over algorithmic risk assessments, it is perhaps not surprising that much of the scholarship in this area focuses on normative and technical assessments of these tools, or on recommendations for how to improve them. With one notable exception (Garrett and Monahan, 2020), scholars have paid little attention to how courts and other criminal justice system actors think about and use these tools.

This is an important gap (Stevenson, 2018). Judges' and other criminal justice actors' attitudes towards and implementation of algorithmic risk assessment tools profoundly affect the role these tools play in decision-making, and, in turn, their impact on defendants, incarceration rates, and the broader criminal justice system (Stevenson, 2018). For example, a judge who trusts risk assessment tools will generally give more weight to a defendant's risk score than one who does not. Additionally, a risk assessment tool will generally perform better when used for its intended purpose than when used for an unintended, "off-label" purpose (Chanenson and Hyatt, 2016, p.7-9). For these and other reasons, "the impacts of risk assessment depend on the same good-old-fashioned factors that have helped and hindered reform for centuries: context, incentives, and details of implementation" (Stevenson, 2018, p. 306).

An understanding of how risk assessment tools function on the ground also matters in another way. As mentioned above, scholars, think tanks, and others have identified a number of best practices for the design and use of risk assessment tools. If we are to assess whether courts' current practices align with these best practices, and so identify ways that courts can improve

their use of risk assessment tools, we must first understand how courts are using these tools today. Further study of how judges and other criminal justice actors think about and implement algorithmic risk assessments is vital to policy discussions about these tools (Chanenson and Hyatt, 2016; Chohlas-Wood, 2020; Garrett and Monahan, 2020).

Scholars have not paid sufficient attention to the judicial attitudes, policies and practices surrounding algorithmic risk assessments (Garrett and Monahan, 2020; Stevenson, 2018). Most empirical work has focused on whether algorithmic risk assessment tools have reduced pre-trial detention and incarceration rates (Berk, 2021; Stevenson, 2018). This is a vital question. But it does not provide direct insight into how courts are utilizing risk assessments or whether they are following best practices. In 2016, Chanenson and Hyatt surveyed a “non-representative and small” sample of judges across multiple jurisdictions to assess their general attitudes towards risk assessments. They found that judicial use of risk assessments during sentencing was “relatively commonplace” and that most, but not all, judges viewed them positively. They concluded with a call for more “research and policy on at-sentencing risk assessment” (Chanenson and Hyatt, 2016).

In 2017 and 2018, Garrett and Monahan surveyed Virginia Circuit Court judges regarding their views on the use of risk assessment in sentencing those convicted of non-violent drug and property crimes (Garrett and Monahan, 2020). Most respondents said that they should consider risk when sentencing such offenders, and that they were familiar with the risk assessment tool for doing so. They further responded that they would issue more non-carceral sentences if their jurisdiction invested more in alternative interventions, and if it required them to state in writing why they sentenced “low-risk” defendants to prison. These findings begin to shed light on judicial attitudes towards and implementation of algorithmic risk assessment tools, but leave unanswered many important questions about how judges utilize these tools and whether they follow best practices.

This report begins to fill this gap. It focuses on the Ohio Courts of Common Pleas—trial courts that exist in each of the 88 Ohio counties, that have original jurisdiction over all state felony crimes, and that are the principal users of algorithmic risk assessments. As is further described below, the researchers surveyed Ohio Court of Common Pleas judges, probation officers, and court administrators regarding their views on and use of algorithmic risk assessment tools.

The team further interviewed judges who volunteered for such interviews, and a diverse array of stakeholders that included victim’s rights, civil liberties and civil rights groups, as well as public defenders and county prosecutors.

The report provides an in-depth picture of how the Ohio Courts of Common Pleas think about and use algorithmic risk assessments. It compares Ohio practice to best practices and, on this basis, formulates recommendations as to how Ohio can best implement risk assessments. By taking an in-depth look at one state, the report also increases knowledge on how criminal justice systems implement risk assessments today and so should inform policy discussions about such tools more generally. The report does not analyze whether algorithmic assessment tools reduce incarceration, or whether they do so in a fair and unbiased way, and so does not provide a basis for opining on the ultimate value of risk assessment tools. Instead, this report focuses on the under-studied questions about how courts view and implement risk assessment tools, questions that shape how these tools impact defendants and the broader society.

2. Methodology

2.1 Study Design

To conduct the study, we distributed the survey to all judges in the Ohio Courts of Common Pleas. After several preliminary interviews with the Ohio Court of Common Pleas and ODRC staff members, we developed questionnaires for both judges and other court staff who play an important role in the risk assessment process (specifically court administrators/deputy court administrators and chief probation officers). We partnered with the Ohio Criminal Sentencing Commission (OCSC), which distributed the surveys to all 242 Ohio Common Pleas judges and with the Ohio Association for Court Administration, which distributed it to court administrators in each of Ohio’s 88 counties. All survey questions were entered into the Qualtrics Survey platform.

We utilized online survey questionnaires to obtain judges’ views and use of risk assessment tools. We chose this type of format as it would allow judges to complete the questionnaire at their own time and pace. This approach is cost-effective and enables the survey instrument to reach numerous respondents. However, it can result in low response rates and limited detailed

feedback. Each participant was given a written consent form and had the option to withdraw from the study at any point or decline participation. Out of the 242 judges contacted, we received 48 responses, which is about a 20% response rate.² Considering that the use of risk assessment tools for sentencing is mandated in Ohio, this low response rate may suggest that many judges did not have time to fill out a 25-minute survey, rather than a lack of familiarity or knowledge of the tools. We recognize that these are low response rates, and we partially circumvent this issue by analyzing the data with non-parametric tests that are slightly better suited for small sample sizes.

To gain deeper insights on judges' opinions and impressions of risk assessment tools, we extended invitations for follow-up interviews to those who expressed a willingness to participate.³ A total of 19 interviews were conducted in the period following the survey closure. The interviewees included 9 judges, 2 court administrators, 2 chief probation officers, and 6 stakeholders, including attorneys (public defense and county prosecutors), civil rights and Civil Liberties Association, ODRC staff members, and academics.

We further recognize the potential for biased findings in our research design due to the non-random selection of participants. The judges who chose to respond to the survey, and those who volunteered to be interviewed, may differ significantly from the broader population. Nonetheless, we maintain that the insights gained from this research can offer valuable perspectives on how some Ohio judges are using risk assessment tools, which can inform further research in this area. All survey responses were fully de-identified. On October 11, 2022, The Ohio State University Institutional Review Board approved the research design.

2.2 Respondent Demographics

Judicial survey respondents were predominantly white/Caucasian and male and ranged from 34 to 71 years old. They came from 29 distinct counties of various sizes and demographics, including the largest county in the state and some of the smallest counties in the state (see Table 1).

²We also surveyed Chief Probation Officers (13 out of 58 for 22% response rate) and Court Administrators (17 out of 181 for 9.4% response rate).

³Although this report mainly focuses on judges, we also interviewed other individuals (9 judges, 2 court administrators, 2 chief probation officers, 2 ODRC staff members, and 1 University of Cincinnati researcher). Additionally, we interviewed several stakeholders about the use of risk assessments in Ohio courts. These included individuals from victim's rights and civil liberties/rights groups, and attorneys (both public defender and county prosecutor).

Table 1. Demographics of judicial survey respondents

| | No. of Respondents⁴ | Mean | Std. Dev. | Min | Max |
|---------------|---------------------------------------|------------------------|------------------|-------------------------------|----------------|
| <i>Age</i> | 44 | 56.75 | 7.95 | 36 | 71 |
| | No. of Respondents | Males | Percent | Females | Percent |
| <i>Gender</i> | 43 | 34 | 79.1% | 9 | 20.9% |
| | No. of Respondents | White/Caucasian | Percent | Black/African American | Percent |
| <i>Race</i> | 44 | 42 | 95.5 % | 2 | 4.6% |
| | No. of Respondents | Mean Population | Minimum | Maximum | |
| <i>County</i> | 29 | 42, 209.48 | 27, 698 | 1, 316, 756 | |

Note: The data represents summary statistics for judges in our sample. The number of respondents in each category varies due to missing values or respondents selecting the option "Prefer not to answer." *Source:* Judge Survey

3. Ohio use of ARAs

3.1 Ohio's adoption of risk assessments

In 2006, ODRC hired the University of Cincinnati (UC) to develop and validate a tool to evaluate the risk of individuals at different points in their journey through the criminal justice system (Latessa et al., 2009). ODRC's motivation was to identify ways to allocate their resources based on the risks and needs of defendants and offenders in a more efficient manner, while also reducing the likelihood of recidivism. With this goal in mind, the Center for Criminal Justice at UC went through a process of planning, data collection, and validation. This process extended for almost three years and involved interviews with more than 1,800 individuals at different points of their criminal justice process, in different counties across the state of Ohio.

When ODRC's and UC's collaboration started, algorithmic risk assessments were not new to Ohio and other jurisdictions. Other tools included the Arnold Ventures' Public Safety Assessment (PSA); COMPAS – Correctional Offender Management Profiling for Alternative Sanctions; LSI-R – Level of Service Inventory – Revised; and the Wisconsin State Risk/Needs Assessment Instrument. These last two tools, LSI-R and the Wisconsin Risk/Needs instrument, were used to compare the

predictive power of the newly developed Ohio Risk Assessment System (ORAS). According to [Latessa et al. \(2009\)](#), ORAS matched or outperformed the other tools.

The results of the [Latessa et al. \(2009\)](#) study informed ODRC's decision to choose ORAS to be used in the state of Ohio, after the legislature had tasked them with the selection of a single validated risk assessment tool for adult offenders in 2011 (ORC §5120.114, 2011). Following ODRC's designation, courts, probation departments, and other entities across the different counties in Ohio began adopting the multiple tools included in the ORAS suite. These include instruments for pretrial, community supervision, institutional/prison intake, and community reentry.

Although stakeholder engagement is recommended during the development of risk assessment tools ([Hamilton, 2020](#)), the ORAS final report by [Latessa et al. \(2009\)](#) does not indicate extensive engagement with community stakeholders. This was confirmed by our interviews:

There would have been conversations with folks that ran programming in the community, but not just from, like, you know, other community perspectives. I'm sure that was not part of the original design (Interviewee No. 19).

It is worth noting that the legislative process to adopt the ORAS tool may have included more broad stakeholder engagement:

[B]ack in 2010/2011, the whole criminal justice reform coalition, which would include all of our community based correctional folks... and you know the conservative think tanks... and the liberal think tanks, you know, all of them agreed that this was a great idea. All of them had a great confidence in a good old Ed Latessa (Interviewee No. 22).

Even though the mandate indicated ORAS should be used when a particular court orders an assessment of an offender for sentencing or another purpose, the implementation was more homogeneous for the instruments used for sentencing, in particular for the ORAS-CST (Community Supervision Assessment tool). In 2019, a Task Force examining the Ohio Bail System, created by the Supreme Court of Ohio, recommended requiring the use of a validated risk assessment tools for setting bond or conditions of bond ([Task Force, 2019](#)). However, the Ohio Supreme Court decided

to not include this recommendation in the package of changes that were proposed to the general assembly (OSBF, 2020).

3.2 Ohio Risk Assessment Process

ORAS assessments involve a comprehensive review of file information, face-to-face interviews, and self-report questionnaires (Latessa et al., 2018). When necessary, additional information is also gathered to corroborate the review and ensure accuracy. Once all data about the offender has been collected, the assessor consults the tool's scoring guide to determine how each item should be scored (see Latessa et al., 2018).

Prior to the administration of risk assessments, the University of Cincinnati Corrections Institute (UCCI) recommends that all ORAS end-users, including court staff and probation officers, attend a two-day on-site training session. The training covers the principles of effective classification (i.e., the risk, need, responsivity, and professional discretion principles). These principles enable agencies to identify who to target to reduce risk of recidivism, what factors or needs to focus on to change criminal behavior, how treatment should be delivered to improve success, and finally professional discretion which empowers agencies to override assessment instruments under unique circumstances. Furthermore, the training includes an introduction to the instruments and a detailed review of how each item on the tools is scored. Participants have multiple opportunities to practice scoring the instruments. Upon completing the training, staff must pass a certification test to ensure they understand how to accurately score the tools. In many cases, judges do not directly conduct risk assessments, so they are not obligated to undergo training and can instead rely on PSI reports prepared by their court staff. No official policy or guideline requires the training of judges.

Each assessment tool in the ORAS suite is designed for specific purposes. For instance, the Misdemeanor Assessment Tool (MAT) is used to evaluate the risk and needs of individuals convicted of misdemeanors, the Prison Intake Tool (PIT) is used to assess individuals upon their entry into prison to determine initial risk levels and needs, the Reentry Tool (RT) is used to assess the risk and needs of individuals nearing release to aid in reentry planning, and the Supplemental Reentry

Tool (SRT) provides a more detailed assessment for individuals with specific reentry challenges. In this report, we focus on the use of the Pretrial Assessment Tool (PAT) and the Community Supervision Tool (CST). PAT is administered at entry into the criminal justice system to assess the offender's flight risk and recidivism risk before trial. As a result, it is primarily used in setting bail. CST is used prior to sentencing and primarily before probation or community supervision and/or during parole or post-release supervision. CST is used to determine the appropriate supervision level and to guide case management for offenders within the community. PAT consists of 7 risk variables, whereas CST consists of 35 risk variables.

3.21 Items on PAT and CST

In the following two paragraphs, we draw upon details from a table summarizing ORAS (see Table 7.1 in [Latessa et al., 2018](#)) and the assessment scoring forms (see Appendix A in [Latessa et al., 2009](#)) to describe PAT and CST. We include examples of specific questions within each assessment and the domains they cover.

The pretrial risk assessment tool is conducted by court staff and is designed to be brief, taking just 10–15 minutes to administer. To accurately assess an offender's risk, several questions are asked across four domains: 1) Criminal History, 2) Employment, 3) Substance Abuse, and 4) Residential Stability. Examples of these questions include the age of the defendant at their first arrest or conviction, the number of failure-to-appear warrants filed in the previous 24 months, whether the defendant was employed at the time of arrest, and whether the defendant had lived at the same residence for the previous six months, among others. Each item is given a raw score and the aggregate composite score is calculated. Total scores range from 0-9 with scores ranging from 0-2 classified as low risk, 3-5 as moderate risk, and above 6 as high risk. Each of these score ranges correlates with percentages that indicate the likelihood of a failure-to-appear or the probability of a new arrest. For example, low-risk offenders (i.e., 0-2 raw PAT score) have approximately a 5% chance of getting re-arrested or failure-to-appear, whereas high-risk offenders (i.e., above 6 raw PAT score) have approximately a 30% chance of re-offending or failure-to-appear.

The community supervision tool is conducted by probation officers and is a much longer as-

assessment tool and takes about 30–45 minutes to administer. To assess the offender’s risk, 35 items are evaluated across seven domains: 1) Criminal History, 2) Education, Employment, and Financial Situation, 3) Family and Social Support, 4) Neighborhood Problems, 5) Substance Abuse, 6) Peer Associations, and 7) Criminal Attitudes and Behavioral Patterns. Examples of the evaluated items or questions include whether a defendant had committed a serious offense under the age of 18, the number of prior adult felony convictions, whether community supervision had been revoked before, employment at the time of arrest, if a defendant was ever suspended or expelled from school, parents’ criminal record, whether a defendant lives in a high crime area, stability of residence, ever used illegal drugs, contact with criminal peers, if offender engages in risk-taking behavior, among others. The total CST score ranges from 0-49. Under CST, there are separate cutoff scores for males and females to avoid over-classification of female offenders. For instance, for male offenders, scores ranging between 15-23 indicate moderate risk, whereas scores between 15-21 indicate moderate risk for female offenders. Again, each of the score ranges correlates with percentages that indicate the likelihood of overall recidivism.⁴

Once the scores are determined, a critical step in the process is to understand how risk assessment scores are communicated to judges. From the literature, it appears that it is the responsibility of the probation officer (or court staff) who conducted the risk assessment to verify, evaluate, and interpret the information gathered, and to present the information to the court in an organized, objective pre-sentence report (see [United States Probation, nd](#)).⁵ This report includes a numeric risk score and other important qualitative information about the offender. The report is disclosed to the defendant, defense attorney, and prosecutor, who can all raise objections through a formal process. The transmission of this report may occur through a secure digital system or in person, ensuring that the information is promptly and accurately delivered. Although practices may differ by jurisdiction, to ensure the judge receives and correctly interprets the score, there are established guidelines and policies. These policies require that the assessment report includes a detailed explanation of the scoring system, the meaning of the scores, and any associated risk

⁴See the scoring forms for specific items – <https://cjdata.tooltrack.org/sites/default/files/2018-10/ORAS%20Scoring%20Assessment.pdf>

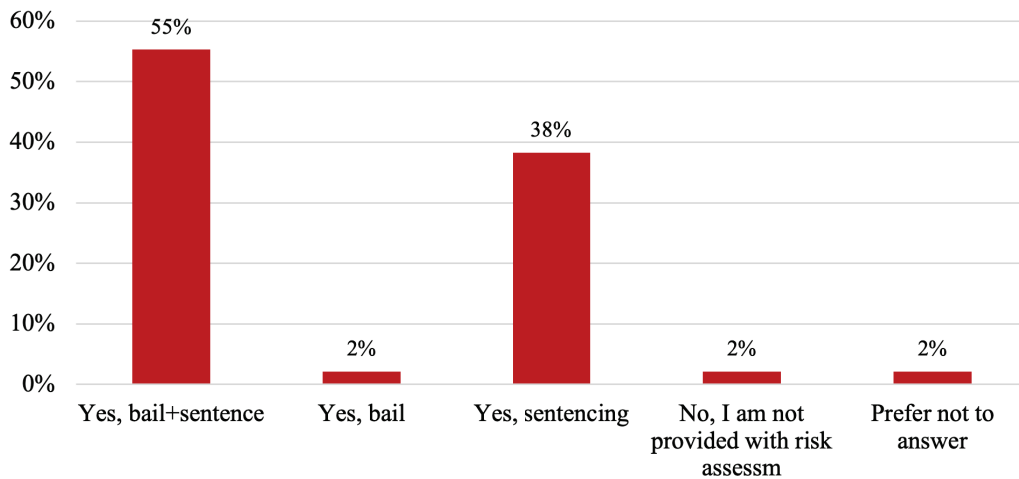
⁵United States Probation - Southern District of Ohio. Retrieved from <https://www.ohsp.uscourts.gov/presentation-investigation>

levels in probabilities or percentages. Additionally, the probation officer or court staff conducting the assessment often provides a verbal briefing or is available to answer any questions the judge may have regarding the report. It is crucial that judges fully understand the implications of the risk scores. Therefore, regular training sessions are conducted for judges to familiarize them with the risk assessment tools and their interpretation. These sessions ensure that judges can make informed decisions based on the assessment reports they receive.

3.22 Risk Assessments in the Ohio Courts of Common Pleas

Most judges use risks assessments for their sentencing decisions (93%). The use of these tools for bail is considerably lower (57%) (see Figure 1). These results are expected as the use of ARA tools is mandated for the case of sentencing, but not for bail, which explains the greater variation in the use of these tools in the pretrial instance.

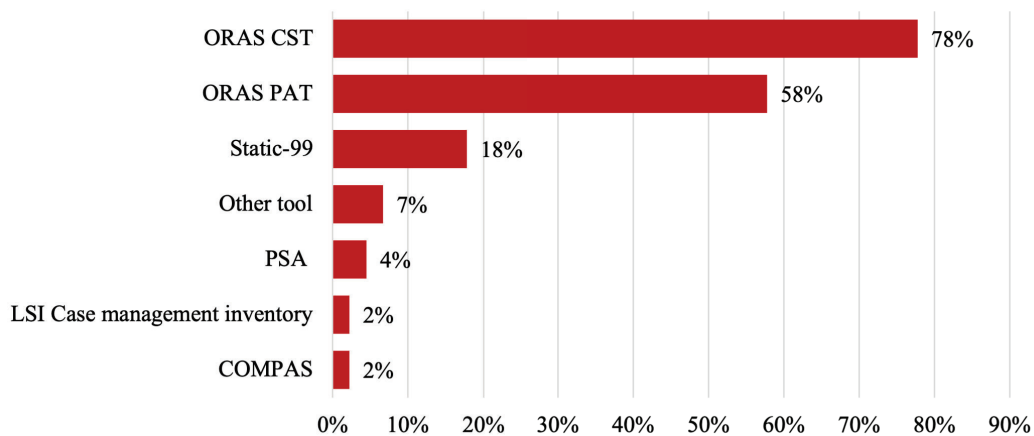
Figure 1. Are you currently provided with ARAs for bail and sentencing decisions?



Note: The data indicate percentage of judges who are provided with risk assessment tools for either bail, sentencing, or both. *Source:* Judge Survey

Most courts use the Ohio Risk Assessment System (ORAS) as their primary tool for bail and sentencing decisions. While a few other tools were also mentioned (see Figure 2 for other tools),⁶ ORAS was far and away the primary tool used by Ohio Courts of Common Pleas in our study, which aligns with prior findings regarding ARA tool use at the pretrial stage in Ohio (Ohio Criminal Sentencing Commission, 2020). This is unsurprising, as Ohio’s Revised Code was updated in 2011 to direct the Ohio Department of Rehabilitation and Corrections (ODRC) to “select a single validated risk assessment tool for adult offenders” across various courts (ORC §5120.114, 2011). Given the lack of significant representation of other tools in our data, we focused exclusively on ORAS, the tool specified by ODRC. As expected, judges using risk assessments for sentencing reported using ORAS-CST, the designated tool, while those applying assessments for bail primarily reported using ORAS-PAT.⁷

Figure 2. Respondents use of tools



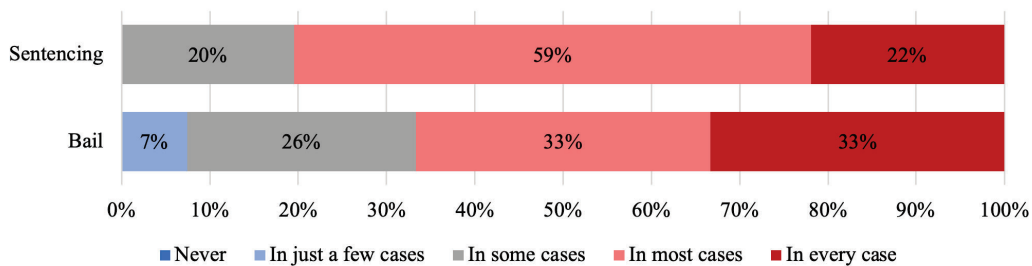
Note: The data indicate the percentage of judges who reported using specific risk assessment tools. *Source:* Judge Survey

⁶This figure includes only tools currently used by judges. The percentages represent the proportion of use among judges in our sample. The following tools are excluded from this figure because none of the judges in our sample reported using them: *Level of Service Inventory – Revised (LSI-R)*, *Level of Service/Risk, Need, Responsivity (LS/RNR)*, *Static Risk and Offender Needs Guide*, *Wisconsin State Risk Assessment Instrument*, and *Virginia Pretrial Risk Assessment Instrument*.

⁷The relationship between reporting the use of ORAS CST and ORAS PAT for sentencing and their use for bail decisions was tested using Chi-squared tests ($\chi^2 = 3.08, \rho = 0.07$ and $\chi^2 = 29.89, \rho = 0.00$, respectively).

In terms of how often judges receive risk assessment scores, they reported receiving ARA scores in most cases, but not all, for bail and sentencing decisions (see Figure 3). For bail decisions, ARA scores may be missed since magistrates often handle these cases—65% of judges noted that magistrates make bail determinations in their courts.

Figure 3. How often do you receive a risk assessment for bail and sentencing?



Note: The data indicate how often judges receive a risk assessment for bail and sentencing decisions. *Source:* Judge Survey

We also learned through our interviews that, in some cases, there may be insufficient staff and/or resources to complete pre-trial risk assessments with all defendants in the short timeframe prior to arraignment where a bail decision is set. When asked about this possibility, one chief probation officer responded as follows:

My guess is, it happens, I would be more than surprised if it’s not a frequent occurrence throughout the state. In my department a number of years ago, once we identified that we had an officer that was from time to time doing that... they got disciplined, the word got around the department that they got disciplined, and we’ve not had it happen since that we know of. But that being said, our caseloads are, I believe, manageable sizes. Now, if you are sitting in a department, maybe that you have 200 offenders on your caseload, and you really feel as though you don’t have the time to dedicate to a 45 minute to an hour session... And like I said, I’d be surprised if it’s not happening frequently across the state. But I don’t know that for fact (Interviewee No. 20).

For sentencing decisions, a few judges mentioned scenarios in which both the defense and prosecution may agree to waive the pre-sentence investigation (PSI) report for a given case. In

virtually all of the courts with which we spoke to judges or staff, the ORAS-CST tool was provided as part of the PSI report. Thus, in those scenarios where a PSI report may have been waived, an ARA may also therefore not be provided to the judge. However, judges noted that this occurrence was rare:

[A pre-sentence investigation report] wouldn't be waived very often. Sometimes it would be...if there's a mandatory sentence, for example, or if an agreement on a plea is reached, just at a different time. Customarily, though, I would guess that in 90 percent of the cases there's going to be a pre-sentence investigation report... maybe a little more. (Interviewee No. 16).

These findings align with expectations for sentencing and bail. For sentencing, where ODRC mandates ARA use, judges consistently receive reports, with 81% reporting ARAs in most or all cases. In contrast, bail decisions show more variation, as ARA use isn't mandated, leaving counties to apply the tools as needed based on available resources.

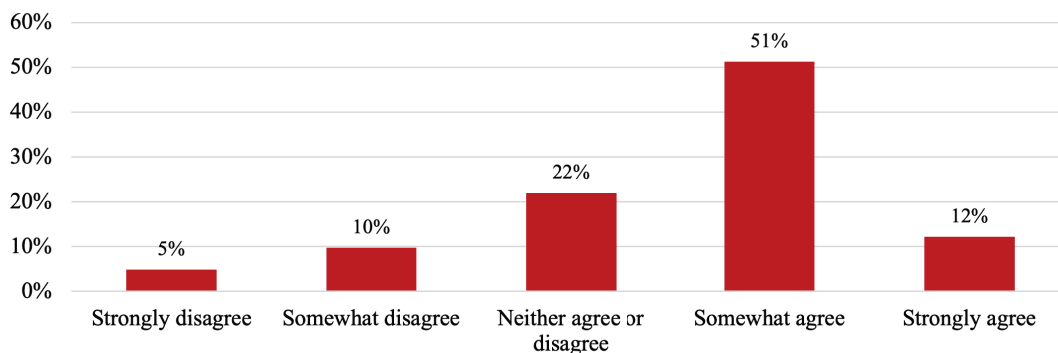
Interestingly, in the case of bail, we see a positive association between how frequently judges receive the ARA, and how important they indicate the assessment is in their decision-making.⁸ This suggests that judges place greater importance on ARAs when they receive them more frequently. Frequent exposure can build judges' familiarity with the tool, increase confidence in its reliability, and reflect the court system's emphasis on ARA use. Conversely, judges who receive ARAs less often may rely on alternative methods and thus place less weight on the assessments. Further research is needed to clarify this relationship.

The survey also asked judges to what extent they agreed with the statement that risk assessments frequently inform their decision-making. In alignment with the results above, judges generally agreed with this statement, indicating they use risk assessments frequently in their decision-making process (see Figure 4).

3.23 Variation in use of ARAs across and within counties

⁸This relationship was tested using a Spearman analysis ($Rho = 0.50, \rho = 0.01$).

Figure 4. Extent judges agree with the statement: “ARAs frequently inform my decision-making.”



Note: The data indicate the extent to which judges rely on risk assessment tools in their decision-making process. *Source:* Judge Survey

When we look at the use of risk assessment tools among judges from different counties, we observe that the variation primarily stems from differences between the use of risk assessment tools for bail. The figure below indicates that, out of the 27 counties for which we have data, about 40% of the counties do not use risk assessment tools for bail and about 60% use ARAs for bail. This indicates an uneven application of the tools across counties for bail decisions. On the other hand, almost all counties indicate using risk assessment tools for sentencing decisions. Only 4% of the counties do not use risk assessment tools for sentencing decisions (see Figure 5). This shows that across counties, there is more consistent application of risk assessment tools for sentencing decisions. This makes sense as the use of ARA tools is mandated for sentencing but not for bail. (ORC §5120.114, 2011).

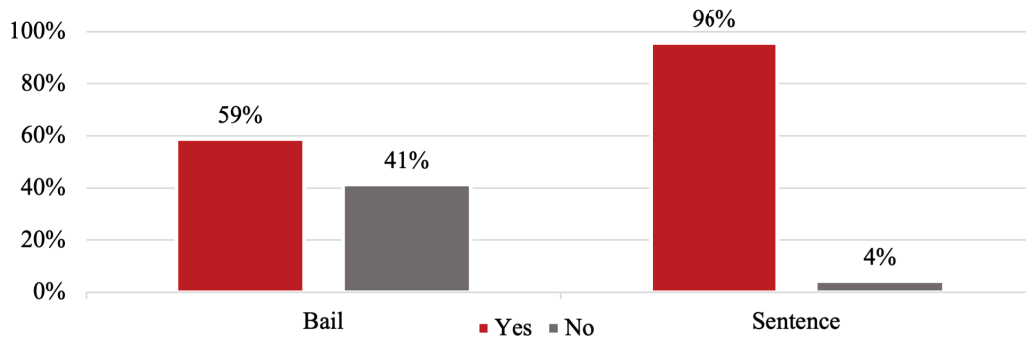
We also noted that large counties, those with a population of more than 100,000, were more likely to use risk assessment tools for bail decisions.⁹ This could be explained due to case volume and resource allocation. In large counties, where caseloads are typically higher, employing risk assessment tools for bail decisions can significantly streamline pretrial decision-making.

Not surprisingly, the use of risk assessment tools for bail is negatively correlated with the percentage of rural population. In other words, counties with a higher proportion of rural residents are less likely to use risk assessment tools for bail decisions. This may be due to logistical challenges in rural areas, such as limited access to technology or limited resources for pretrial services.

⁹We tested this relationship using Pearson’s Chi-squared test ($\chi^2 = 8.20, \rho = 0.00$).

Additionally, rural areas often have fewer personnel available for managing pretrial processes. This may make it difficult for rural courts to dedicate time to conduct individualized risk assessments of defendants.

Figure 5. Variation in risk assessment usage across counties



Note: The data indicate the variation in the use of risk assessment tools across different counties. The graph compares the use of tools for bail and sentencing decisions separately. The percentages indicate the proportion of counties in our sample using or not using risk assessment tools for bail and sentencing decisions *Source: Judge Survey*

Looking at the responses judges from the same counties gave us to this question, we also found some variation within jurisdiction. Among the eight counties from which we observe responses from multiple judges, only three have consistent responses regarding the use of risk assessment tools. Consistent responses within counties indicate that all judges' responses are the same, suggesting that the use or application of risk assessment tools align (i.e., they all select using ARAs for sentencing only, etc.). However, judges from the other five counties report inconsistent use. For example, within one county, we notice that one judge use ARAs for bail and sentencing while the other only uses them for sentencing. These results could indicate differences in the use of ARA tools by different judges in the same jurisdiction, gaps in the knowledge judges have regarding the tools that are used in their county, or both simultaneously.

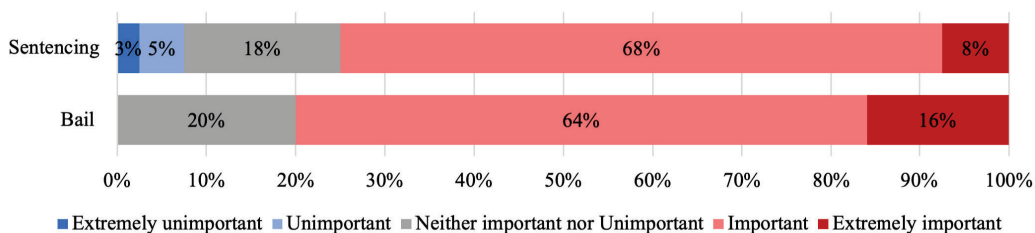
3.3 Judicial views and attitudes towards ARAs tools

In addition to how judges use risk assessment tools in practice, our survey asked several questions about judges' views and attitudes towards risk assessments. This information is important since such views and attitudes may influence the ways judges use ARA tools in their courts.

3.31 Overall importance of ARAs for decision-making

As shown in Figure 6, judges generally indicated that they find ARAs important to both bail and sentencing decisions, with very few considering them unimportant.

Figure 6. How important do you consider risk assessments to your bail and sentencing decisions?

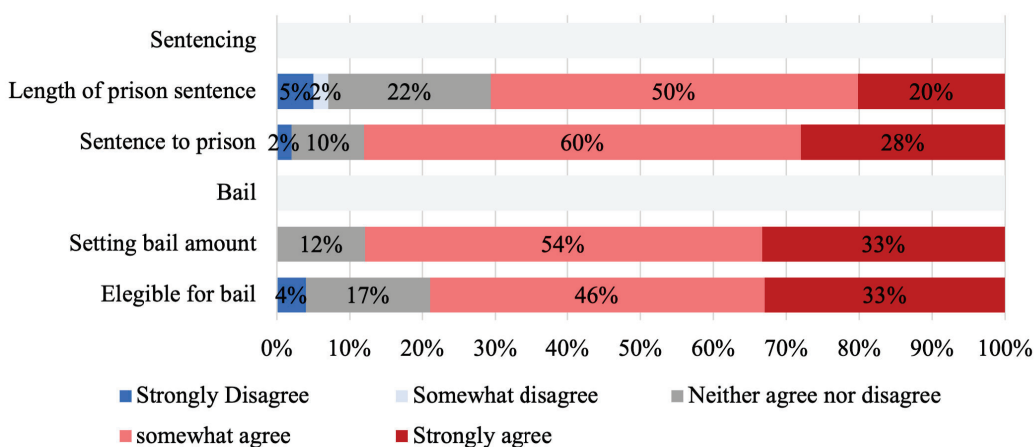


Note: The data indicate the importance of risk assessment tools in informing bail and sentencing decisions. *Source:* Judge Survey

We also asked respondents how much they believe judges should rely on risk assessment scores during bail and sentencing decisions. For bail, we inquired if courts should consider risk assessments

when deciding eligibility for bail and setting the amount. Regarding sentencing, we explored whether judges should use risk scores to decide on imprisonment and the length of sentences. Overall, judges' views on risk assessments were consistent across scenarios, but they valued these assessments more for deciding on imprisonment than for determining sentence length.¹⁰ Overall, respondents either agreed or strongly agreed that risk assessments were valuable in each specific bail and sentencing scenario (see Figure 7), a response that largely correlates with judges' views on the ARA's importance for bail and sentencing decisions more generally (see Figure 6).¹¹

Figure 7. Relevance of ARAs for specific decisions



Note: The data illustrate the relevance of risk assessment tools for specific decisions. *Source:* Judge Survey

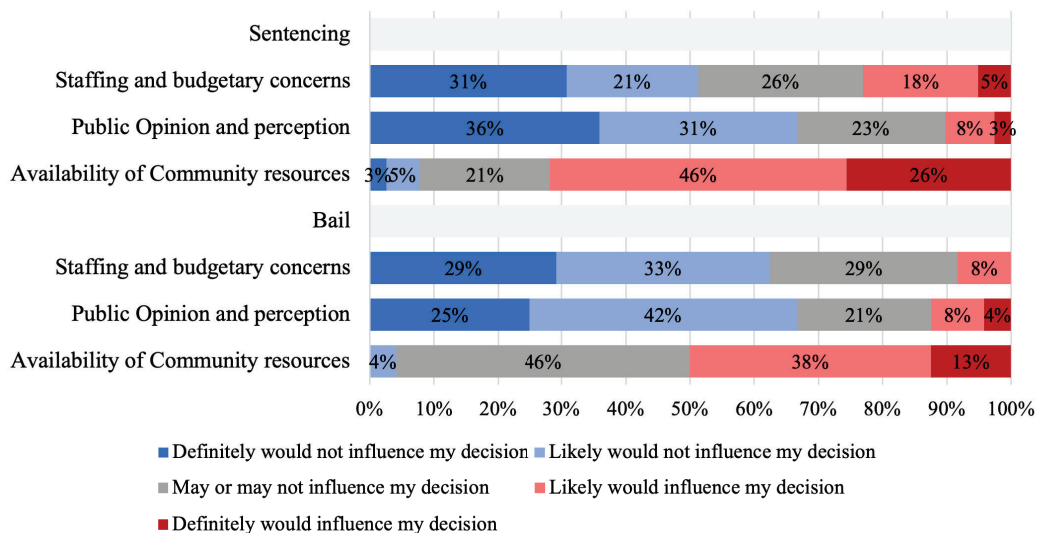
¹⁰Using a Wilcoxon signed-rank test, we determined there is no variation in how judges assess ARAs for determining eligibility and amount in bail ($z = 0.744, \rho = 0.68$), and that there is a significant difference in how they evaluate these tools for both sentencing dimensions ($z = 2.825, \rho = 0.00$).

¹¹These relationships were tested using a Spearman correlation. In the case of bail, we found a statistically significant relationship for determining the eligibility and for setting the amount ($Rho = 0.48, \rho < 0.05$, and $Rho = 0.55, \rho < 0.01$, respectively). Using the same test for sentencing, the importance of ARAs and its particular use to decide whether to sentence someone to prison were significantly correlated ($Rho = 0.60, \rho = 0.00$). In contrast, the relationship between the importance of ARA for sentencing, and for determining the length of prison sentence was not statistically significant.

The survey asked respondents to indicate what factors can influence the weight they give the ARA score for bail and sentencing decisions (see Figure 8). In both cases most judges agreed that the availability of community resources for those released was the more highly rated factor with 51% for bail, and 72% for sentencing, indicating that the factor would influence their decision to some extent. These results align with those in [Garrett and Monahan’s 2020 study](#), indicating that judges were more likely to override the results of ARAs when the availability of community resources was limited.

In comparison with the other factors listed, respondents also agreed that the availability of community resources is more important than both staffing and budgetary concerns and public opinion.¹² The intensity of these opinions differs depending on the age and experience of the judge. On one hand, older judges tend to give less importance to staffing and budgetary concern than younger judges, while more experienced judges give less importance to public opinion.¹³

Figure 8. Factors that influence weight judges give to risk scores



Note: The data show several factors that may influence judges’ decisions in the use of risk assessment tools. *Source:* Judge Survey

¹²This was tested using a Mann-Whitney test with results significant at 90%. In the case of bail, 75% think community resources are more important than public opinion, and 87.5% think it is more important than staffing and budgetary concerns. In the case of sentencing, 76.6% of judges think community resources are more important than public opinion and 66.6% think it is more important than staffing and budget.

¹³These relationships were tested using the Spearman correlation test, with both results significant at 90%: staffing and budget (bail $Rho = -0.64$; sentencing $Rho = -0.32$), and public opinion (bail $Rho = 0.34$; sentencing $Rho = -0.31$).

Judges' perceived relevance of availability of community resource for those released was also shared by the stakeholders interviewed. One interviewee highlighted the importance of looking at supportive measure in their relationship with the scores and the way judges would perceive risk:

[M]itigation is key. Risk assessment looks at numbers and data and it's just not looking at how support actual supportive measures could influence those scores or your perceived risk. (Interviewee No. 5).

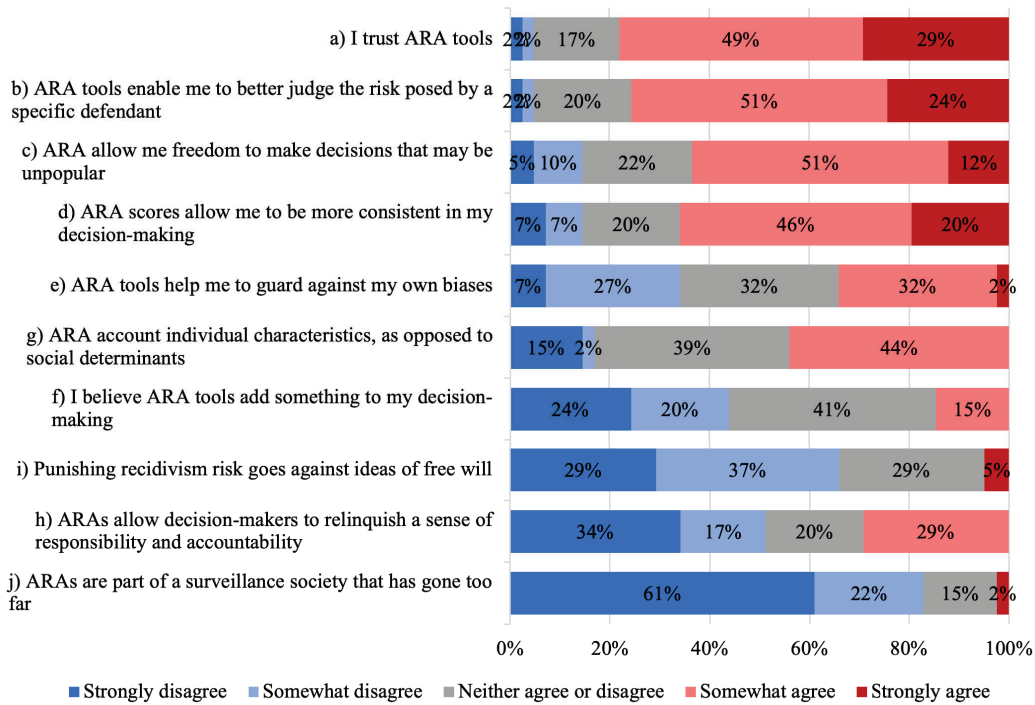
3.32 General views towards risk assessment tools

Several survey questions investigated judges' general views towards risk assessments, and their usefulness for certain aspects of their decision-making. We asked judges to what extent they agreed with the following statements: a) I trust risk assessment tools, b) Risk assessment tools enable me to better judge the risk posed by a specific defendant, c) Risk assessments allow me the freedom to make decisions that may be unpopular, d) Risk assessment scores allow me to be more consistent in my decision-making, e) Risk assessment tools help me to guard against my own potential biases at the point of decision-making, f) I believe risk assessment tools add something to my decision-making process, g) Risk assessment tools account for individual characteristics, as opposed to social determinant, h) Risk assessment tools allow decision makers to relinquish a sense of responsibility and accountability, i) Punishing recidivism goes against ideas of free will central to law, and j) Risk assessments are part of a surveillance society that has gone too far.

Overall, judges generally agreed with these statements (see Figure 9). For instance, most judges strongly agree or somewhat agree that they trust risk assessment tools (78%) and that they enable them to judge better the risk posed by a specific defendant (75%), that ARA tools allow them to make decisions that are unpopular (63%), and that they allow them to be more consistent in their decision-making (66%). In contrast, for other statements, there seems to be less agreement among judges. For example, only 44% of judges agreed with the statement indicating ARAs account for individual characteristics as opposed to social determinants. Similarly, only 34% agreed that ARAs help them guard against their own biases.

We observed that more judges strongly disagree or somewhat disagree with the statements that punishing recidivism risk goes against ideals of free will (66%), that risk assessment tools allow decision makers to relinquish a sense of responsibility and accountability (51%), and that risk assessments are part of a surveillance society that has gone too far (83%).

Figure 9. Judges' views towards ARA tools



Note: The data indicate judges' general views towards risk assessment tools. *Source:* Judge Survey

Most items in Figure 9 are also positively associated with how important judges consider ARA scores in sentencing decisions and how frequently they use these tools.¹⁴ In other words, the more judges trust risk assessments—believing they enhance decision-making, improve consistency, or help reduce bias—the more likely they are to prioritize and regularly use these tools. The strongest link is with the belief that ARAs add value to sentencing decisions, and this is also the strongest correlation for bail decision-making. This suggests that judges may value these tools largely because they provide additional, practical insights for making informed decisions.

¹⁴These relationships were tested using the Spearman correlation test. All the relationships listed here are significant at 95%: Importance for bail: b) $Rho = 0.48$, d) $Rho = 0.41$, e) $Rho = 0.51$, f) $Rho = 0.53$. Importance for Sentencing: a) $Rho = 0.52$, b) $Rho = 0.55$, c) $Rho = 0.48$, d) $Rho = 0.60$, e) $Rho = 0.45$, f) $Rho = 0.70$, g) $Rho = 0.32$, j) $Rho = -0.46$; Frequency of use: a) $Rho = 0.43$, b) $Rho = 0.53$, c) $Rho = 0.37$, d) $Rho = 0.61$, e) $Rho = 0.45$, f) $Rho = 0.66$, g) $Rho = 0.33$, j) $Rho = -0.41$.

For bail decisions, judges tend to assign importance to risk assessment scores regardless of their overall trust in the tool. However, judges who feel that ARAs help them better assess a defendant's risk are more likely to emphasize these scores in both bail and sentencing decisions. Put simply, even if they don't fully trust ARAs in bail contexts, judges still value the insights these tools offer for assessing defendant risk, which drives their importance. Although this may seem counterintuitive, it may reflect the unique aspects of bail decisions, as noted by one of our interviewees (Interviewee No. 3).

[Bail] is the most dangerous decision that I make on a daily basis [...] because [it] is the decision that I make with the least amount of information... I think a lot of times the risk assessment is a check on the things you find out at bail hearing... I'm using it in evaluating my bail decision, I'm sort of using some of the things that they use in the risk assessment that are standardized and [checking] '[O]h wait! Why is that sort of not in line with my thinking? ... how am I looking at this different than the evaluator?

The same judge, when asked whether he trusted ARAs, indicated:

[For bail] I guess I trust [ARAs] less, probably at that point in the proceeding, because I have the [least] amount of information. If I were to sentence somebody or set up bail and have no other information other than a risk assessment, score. That would be terrible, but I also know sometimes you only have that much information...

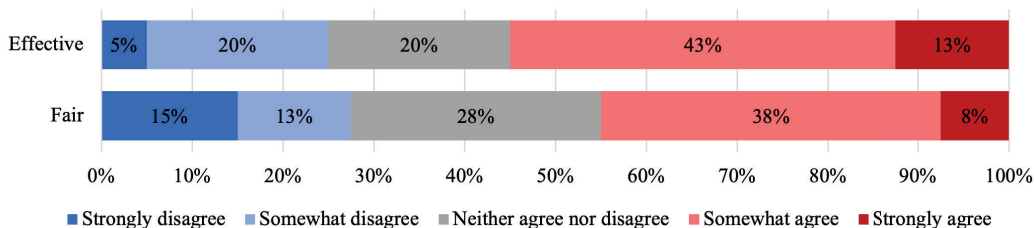
Judges noted that a major challenge in bail decisions is the limited information available to them. Due to this scarcity, they may assign greater importance to ARA scores, even if they're uncertain about fully trusting the scores.

We further observed notable differences in how age, race, and gender shaped judges' views on risk assessment tools. Younger judges, on average, were more likely to believe that ARAs help them mitigate their own biases, expressed greater trust in ARAs, and felt that these tools gave them the flexibility to make unpopular decisions.¹⁵ Additionally, judges eligible for reelection were

more likely to agree that ARAs help them assess defendants' risk more accurately, support them in making unpopular decisions, and adequately account for individual characteristics rather than social determinants.¹⁶

Judges' general views of the judicial system could have an impact on the way they view and the importance they give ARAs in their decision-making. The survey asked respondents whether they think the current bail system is fair or effective, whether the incarceration rates are appropriate or too high, and if treatment resources available in their jurisdictions are adequate. Regarding the cash bail system, judges agree more strongly with the notion that the system is effective (55% agree to some extent) than that it is fair (45%; see Figure 10). These beliefs are correlated with the importance judges give to ARAs for bail decisions. Judges consider the tools to be more important when they think the system is less effective or less fair, in which case they also tend to use the tools more frequently.¹⁷

Figure 10. Judges believe the current cash bail system is:



Note: The data show judges' belief on the effectiveness and fairness of the current cash bail system. *Source:* Judge Survey

In sentencing decisions, judges' views on incarceration rates and local treatment resources do

¹⁵Spearman correlation was used to test the relationships between the different items and age. The relationship with the following items was significant at 95%: a) Risk assessment tools help me to guard against my own potential biases at the point of decision-making ($Rho = -0.36$), b) I trust risk assessment tools ($Rho = -0.32$), c) Risk assessments allow me the freedom to make decisions that may be unpopular ($Rho = -0.44$). The relationship with tenure was also tested with no significant result.

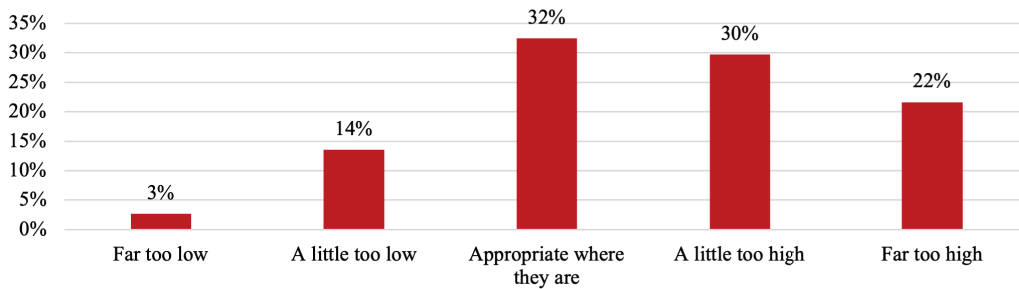
¹⁶The Mann-Whitney test was used to test the relationships between the different items and Reelection eligibility. The relationship with the following items was significant at 95%: a) Risk assessment tools enable me to better judge the risk posed by a specific defendant ($z = -2.59$), b) Risk assessments allow me the freedom to make decisions that may be unpopular ($z = -1.77$), c) Risk assessments sufficiently account for the defendant's individual characteristics, as opposed to social determinants ($z = -2.31$). The relationships with gender and race were also tested with no significant result.

¹⁷Spearman correlation was used to test this relationship with the following relationships significant at 95%: importance and fairness ($Rho = -0.53$), importance and effectiveness ($Rho = -0.59$), frequency of use and effectiveness ($Rho = -0.31$).

not appear to affect how much importance they assign to ARAs or how often they use them. This finding contrasts with Figure 8, which shows that the availability of community resources is the most influential factor in determining the weight given to ARA scores. However, Figure 8 addresses a hypothetical factor’s relevance, not necessarily indicating whether it would increase or decrease reliance on ARAs. The differences between Figure 8 and Figure 12 suggest that while community resources matter universally, their effect on ARA weight may vary—sometimes increasing relevance, other times decreasing it.

In the case of incarceration rates, we see that 52% of respondents think the rates across the judicial system are higher than what they would consider appropriate (see Figure 11), while for availability of treatment, 65% of judges agree to some extent that resources in their jurisdictions are adequate (see Figure 12). These views are related to the age and experience of the judges. Older and more experienced judges tend to think the incarceration rates are not as high as younger and less experienced judges.¹⁸

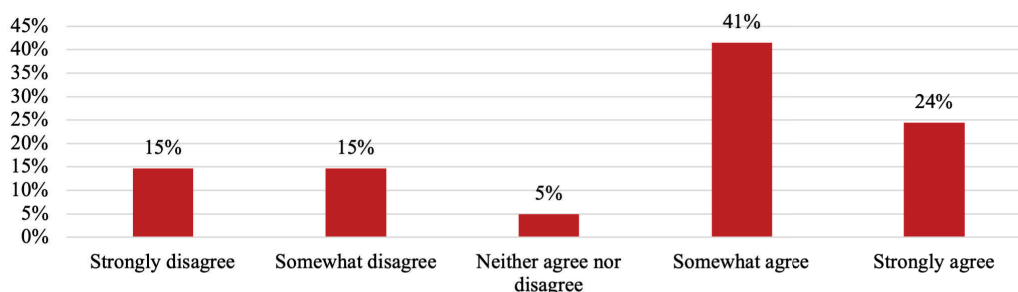
Figure 11. Judges believe the incarceration rates across the judicial system are:



Note: The data show judges’ belief on whether the incarceration rates across the judicial system are higher than what would be considered appropriate. *Source:* Judge Survey

¹⁸The relationship was tested using Spearman correlation, with significant results at 95%: Age ($Rho = -0.44$); Experience ($Rho = -0.41$).

Figure 12. Judges believe the treatment resources in their jurisdiction are adequate.

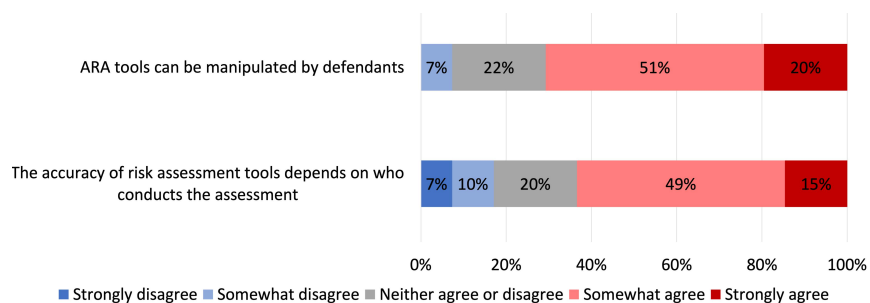


Note: The data show judges' belief on whether the treatment resources in their jurisdiction are adequate. *Source:* Judge Survey

3.33 Views on accuracy

We were interested in some of the factors which may make judges less likely to trust risk scores. To that end, we asked judges whether they believed that defendants could manipulate their risk scores, and whether they thought that the accuracy of risk assessment tools depends on who conducts the assessment. Respondents mostly agreed with both statements (see Figure 13).

Figure 13. Judges' views on the accuracy of the tools



Note: The data reveal judges' opinions on whether the accuracy of risk assessment tools depends on the assessor and whether they can be manipulated by the defendant. *Source:* Judge Survey

These results indicate that judges recognize an element of variability in these tools and do not see them as purely objective or absolute. Rather, they recognize that both the defendant and the person conducting the assessment can potentially influence an ARA score. The acknowledgment that defendants can potentially manipulate risk assessment scores suggests that judges may believe such scores to be biased in defendants' favor. However, the variability in the score based on who conducts the assessment is less clear in terms of what the outcome or implication of such variability might be. For example, a chief probation officer interviewee explained that the person conducting

the risk assessment may take steps to shorten the process in some cases:

[T]he other part to it is, which you know I have witnessed first-hand over the years is, you know, people don't do the ORAS, don't administer the ORAS the way that they were trained and certified to do. Because once you get used to the questions and you, you know, you're looking at someone's background, I mean, I'm like I really don't need to ask that question. I know what the answer is...I know that's a high crime neighborhood. I don't need to ask them that, you know, and so you can kind of check it off. So, you take what maybe [is] designed to be a 45 min to an hour interview, and now it's down to 15 or 10 minutes (Interviewee No. 20).

The interviewee specifically mentioned this scenario in the context of court resources. That is, the suggestion that such shortening of the assessment interview may be done in cases where there are less staff and/or the existing staff already feel overburdened or stretched thin. Again, though, it is unclear precisely how such differences in conducting the assessment might affect the scoring, and therefore the outcomes for defendants.

It should also be noted that some interviewees indicated that they did not feel risk assessment tools were particularly good for evaluating the risk that defendants posed with respect to particular categories of behavior, such as sexual deviancy or substance use:

It is not an accurate picture of those who do have severe substance use; it's not an accurate depiction of exactly how severe that is (Interviewee No. 6).

It is worth noting that the ORAS-CST does include a specific domain for substance abuse, but the above quote was made with that in mind, suggesting the individual found the substance abuse domain insufficient. Other, more specific risk assessment tools designed to evaluate specific behaviors such as sexual deviancy and substance abuse are available. However, not all counties make use of such additional tools, and thus may only have the ORAS tools to base decisions on.

Judges' perceptions of the accuracy of ARAs are linked to their overall views on how these tools improve decision-making. The more judges believe that defendants can manipulate ARA scores, the less they trust the tools, the less they feel ARAs promote consistency in their decisions, and the

less they believe ARAs add value to their decision-making process.¹⁹ Judges' experience appears to influence their perception of ARAs being susceptible to manipulation by defendants, with older judges more likely to agree with this view.²⁰ In contrast, there is no significant relationship between judges' views on the accuracy of ARAs depending on the assessor and their overall opinions of the tools. This suggests that variability in accuracy due to defendants has a much greater impact on how judges think about and use ARAs than variability caused by internal evaluators.

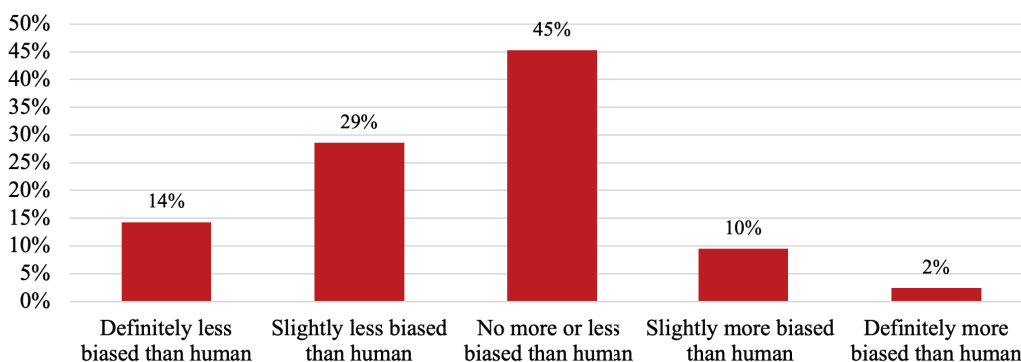
3.34 Judicial views on bias

Another issue we explore is whether the tools themselves are biased, which is a concern that has been raised by many (see [Eckhouse et al., 2019](#); [Mayson, 2018](#)). Our survey asked judges two separate questions that related to the broader issue of bias: 1) How biased they believe risk assessments are compared to human decision-makers, and 2) How they feel risk assessment tools compare to their own ability to assess defendants' level of risk. Although the second question does not directly address bias, it touches on a similar concept. Our results reveal that judges generally perceive ARAs as either equally or less biased than human decision-makers (Figure 13). In contrast, judges tend to view their own judgment as better than, or at least on par with, ARAs (Figure 14).

¹⁹We used a Spearman correlation test for the relationship between opinion defendant manipulate and their level of trust on ARAs ($Rho = -0.36, \rho < 0.05$), ARAs' ability to enable them to make consistent decisions ($Rho = -0.33, \rho < 0.05$), and the extent to which they add something new to the decision-making process ($Rho = -0.33, \rho < 0.05$).

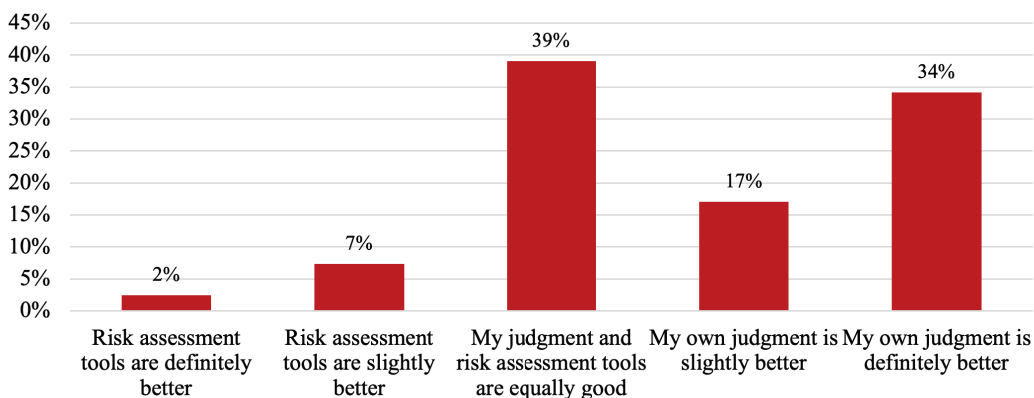
²⁰This relationship was tested using a Spearman correlation ($Rho = 0.33, \rho = 0.03$). Age was also tested using the same tool, but with no significant results. For race, gender, and reelection eligibility we used the Mann-Whitney test, with no significant results.

Figure 14. Judges' views of ARA bias compared to human decision-makers



Note: The data reveal judges' opinions on whether risk assessment tools are more or less biased compared to humans. *Source:* Judge Survey

Figure 15. Judges' responses to the question, "How do you feel ARA tools compared to your own ability to assess defendants' level of risk?"



Note: The data reveal judges' opinions on the ability of risk assessment tools in evaluating risk compared to human judgement. *Source:* Judge Survey

These results suggest that judges may view their own judgment as superior to that of the average human decision-maker. When examining individual responses,²¹ we see that 58% of respondents consider their judgement as better than ARAs, while also considering ARAs to be generally less biased than humans. The remaining 42% of respondents ranked both questions similarly, with no judges rating their own judgment as worse than the average human decision-maker compared to ARAs. This may suggest a self-serving bias, where judges view their own abilities more favorably than those of others. Regardless of the accuracy of their self-assessment, judges' perceptions of both ARAs and their own judgment influence how they use these tools. Interestingly, judges' views

²¹Using a Wilcoxon signed-rank test ($z = 5.11, \rho = 0.00$)

on ARAs being less effective than their own judgment shift with age and experience.²² Older judges and those with more years of experience are more likely to believe their judgment is superior to that of ARAs compared to younger or less experienced judges.

Regarding bias specifically, we also asked judges if they think risk assessment tools discriminate against defendants based on protected class. Overall, judges responded “No” (56.10%) or “I don’t know” (36.59%) to this question rather than “Yes” (7.32%). Additionally, none of those who viewed risk assessment tools as more biased than human decision-makers also believed the tools discriminate against protected classes. This suggests that the bias judges perceive in ARAs is not primarily related to discrimination against defendants from protected classes. Further analysis shows a positive correlation between the perception of bias and the belief that defendants can manipulate the ARA, but no correlation with the idea that the assessor influences the outcome, indicating the bias is seen as stemming from defendant manipulation, not the assessor.²³

It should also be noted that perceptions of bias may impact the importance that judges give to ARA tools, and thus, their use. The extent to which judges believe ARAs to be more biased, relative to human decision makers, is significantly and negatively correlated with their feeling that ARAs are important to both bail and sentencing decisions, as well as how frequently ARAs are used to inform decision-making.²⁴ Thus, judges may weigh ARA scores less in their decision-making process if they feel they are biased relative to human decision makers. Similarly, judges that consider their judgement is better than ARAs tend to give less importance and use ARAs less frequently for decision-making.

Judges’ views contrast with the position expressed by stakeholders. Bias was one of the most salient concerns for our interviewees, as they believe it can be built into the ARA tools by the use

²²Using Spearman correlation, age and years in service had a statistically significant relationship with judges’ perception regarding their own ability compared to ARAs ($Rho = 0.43, \rho = 0.00$, and $Rho = 0.32, \rho = 0.03$, respectively). These variables were also tested against race, gender, and reelection eligibility, with no significant results.

²³Using the Spearman test: defendants can manipulate ARA score ($Rho = .30, \rho = .08$) and the accuracy depends on the individual conducting the assessment ($Rho = .14, \rho = .39$).

²⁴These relationships were tested using the Spearman correlation test. All the relationships listed here are significant at 95%: -ARAs are biased compared to human decision makers and 1) ARAs importance for bail: $Rho = -0.42$, 2) ARAs importance for sentencing: $Rho = -0.58$, 3) ARAs frequently inform decision-making: $Rho = -0.68$, 4) ARAs help guard their own bias: $Rho = -0.66$. -ARAs are worse than the judge’s judgment and 1) ARAs importance for sentencing: $Rho = -0.39$, 2) ARAs frequently inform decision-making: $Rho = -0.38$, 3) ARAs help guard their own bias: $Rho = -0.47$.

of historical data:

I think it's honestly as easy as bias in, bias out...the biggest problem that we have with risk assessment tools is that any algorithm created from criminal justice system data is going to contain racial bias because we know that there is bias in the system all the way from over-policing in certain communities like communities of color all the way to sentencing (Interviewee No. 5).

[T]he concept of having a tool where everything related to race, everything related to sex, everything related to socioeconomic status would be taken out of the equation, and you would purely make a decision based on, you know, how prior decisions have been made, you know, the prior recommendations...and you know it would, there would be some predictability and uniformity in terms of defendants knowing what to expect. The idea of that, I think, is great, and no one would really disagree with that – that there should be predictability and uniformity, especially in terms of things like release conditions and sentencing and probation terms. I think the problem is, the tools are built by humans who have the same bias that they're building into the tool. So, it's not really working the way that it was designed to work, even though the idea of it is, I think, probably not objectionable to anyone (Interviewee No. 11).

Given all of these concerns, most of the stakeholders we spoke to were generally against the use of risk assessment tools. However, some of them did acknowledge the potential usefulness of risk assessment tools, even if recognizing their limitations:

I think, for some courts like I said, it's very beneficial, so I wouldn't say that they should never be used, because I think that there are some people who are [going to] have inherent biases that far exceed what the risk assessment tool would have built in. And so obviously for those people, it's like a lesser of two evils situation where it's better to have this tool that's at least been vetted where at least there's some uniformity as opposed to the Wild West in that courtroom. But I think for a lot of our

judges who are really good, like I said, forcing them to have their decisions be beholden to something that has inherent flaws is problematic where they can do a better job if they're allowed to let more subjectivity enter into it (Interviewee No. 11).

Despite the striking differences in the perception of bias expressed by judges and other stakeholders, the disagreement may not be caused by active opposition. Rather, it may be explained by lack of knowledge, as expressed by the one of the judges interviewed when asked whether they considered the tools to be biased:

That's the whole point, to eliminate bias, I mean. I hope not! Because that's why I rely on them [ARAs]. I mean...that would be horrific if the risk assessment tools have a built-in bias... I have trust that they're not [biased]. I think that they're not and hope that they're not...if I thought that they were [biased] I wouldn't use them (Interviewee No. 4).

3.35 Variation in views across counties

One of the themes that we saw come up consistently in our data collection was the issue of inconsistency in the use of risk assessment tools. Generally, best practices recommend that the use of risk assessment tools should be standardized across jurisdictions and users (Hamilton, 2020). However, Ohio's court system more broadly is built to do precisely the opposite – to allow the different entities at the local level to set most of their own policies. This impacts how common pleas courts set their policies and procedures for several areas, including the use of ARAs. To some extent, this makes the lack of standardization in the implementation of ARAs an inherent feature of Ohio's system.

Regardless of the inability to prescribe unique processes and policies to use ARAs given that the system is structured in a decentralized manner, there are still multiple ways in which consistency may be considered. While ODRC administers training and oversight of ORAS in a generally consistent manner, a variety of factors influence how ORAS tools are used in actual practice by individual courts. One ODRC interviewee indicated that a previous external study described Ohio courts as “a patchwork of systems” (Interviewee No. 17). That individual would go on to say:

I would say that is still true today in that each of the counties, in municipal courts and halfway houses, CBCFs, even ODRC, we each have our own policies regarding how and when to use the tools and procedures of using those tools. So, I would not say that that's consistent across the board of when we even assess, or how and what that's used for. So with that being said, I think that that leaves opportunity for something to make it more consistent so that if it's done one way in a county in the North, the tool means the same thing and is done the same way in [] a county in the South, for instance. . . That makes it also difficult, though, to make any certain specific recommendation about how and when they use it, because everybody uses it based on their own determination by their own administrator (Interviewee No. 17).

The above quote highlights one of the central issues that was raised by multiple interviewees in our study. A court administrator further echoed concerns about consistency between courts:

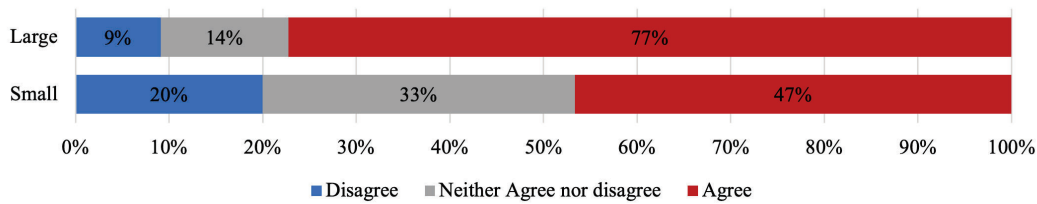
None of the counties are uniform at anything that they do. We don't have any type of unified court management system...Everybody does, you know, kind of the things that they want to do...It depends on the political culture of the county (Interviewee No. 15).

Earlier, we mentioned that judges overall indicated they frequently use ARAs to inform their decision-making, they trust risk assessments, and they believe that risk assessments better help them to judge the risk posed by defendants (Figure 9). However, some interesting relationships were observed when we looked at differences across judges and jurisdictions. For example, we note that statements regarding the frequency with which judges use ARAs to inform decision-making were, on average, endorsed more highly by judges from relatively larger counties (population of more than 100,000)²⁵ than by those from relatively smaller counties (population of less than 100,000;

²⁵We recognize that counties could be categorized in numerous ways regarding size. We chose to use the distinction that we did as it represented a natural break in our data regarding size of county that our respondent judges were based in, and it also allowed for a fairly even split between larger (24) and smaller (17) counties. Ohio's 27 largest counties are above this threshold in population, while 61 are below it. As larger counties have more judges in their courts, there were correspondingly more judges from larger counties in our distribution list, and thus it reasons that our sample would be somewhat skewed towards larger counties as well. We acknowledge that these categories are not objective but are rather described as larger and smaller in relation to each other, and we feel do still approximate such differences across Ohio counties to a reasonable degree.

see Figure 16).²⁶

Figure 16. ARAs frequently inform my decision-making by county size



Note: The data show how often judges' decisions are informed by risk assessment tools in large and small counties. *Source:* Judge Survey

Similarly, judges from urban counties indicated their trust in ARAs is stronger than those in more rural counties (see Figure 17).²⁷ These results are consistent with Stevenson's (2018) findings, indicating judges in urban Kentucky used ARAs more than their peers in rural areas of the same state. Given these results, it is perhaps not surprising that we also saw that judges from smaller counties, and counties with higher percentages of rural population, were more likely to agree with the statement that risk assessments can be manipulated by defendants (Figure 13).²⁸ While we are not able to determine if a belief that defendants can manipulate ARAs leads to lower trust/frequency of use or vice versa, it is clear that these measures are related. Indeed, beliefs that ARA tools can be manipulated by defendants is significantly and negatively correlated with both trust of ARAs and frequency of use.²⁹

Further, we also note that judges from relatively smaller, rural counties on average indicated that they felt the incarceration rates across the judicial system were more acceptable than judges from larger counties, who on average felt they were too high.³⁰ This may potentially provide an explanatory mechanism regarding the differences in trust and frequency of use for ARA tools. That is to say, if judges believe that use of risk assessment tools are more likely to divert defendants from prison, as advocates assert (see [Stevenson, 2018](#)), then a judge's belief about current incarceration

²⁶This relationship was tested using the Mann-Whitney test ($z = -1.93, \rho = 0.05$).

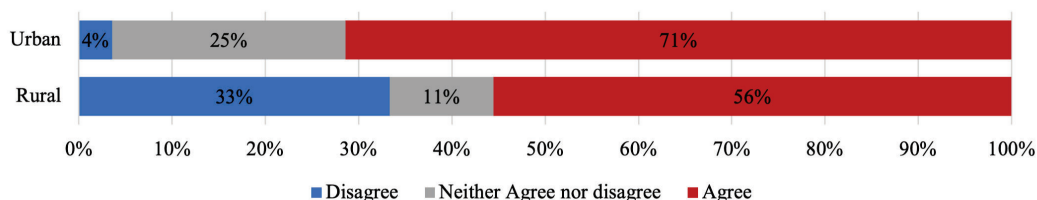
²⁷This relationship was tested using the Mann-Whitney test ($z = -1.75, \rho = 0.09$).

²⁸We used the Mann-Whitney test for large counties ($z = 1.84, \rho = 0.05$), and Spearman correlation for percentage of rural population ($Rho = 0.3, \rho = .07$).

²⁹Tested using Spearman correlation, with trust of ARAs ($Rho = -.52, \rho < .001$), and frequency of use ($Rho = -.212, \rho = .167$).

³⁰Mann-Whitney test for large counties ($z = -2.57, \rho = 0.01$), urban counties ($z = -1.86, \rho = 0.06$).

Figure 17. I trust ARA tools by urban and rural counties

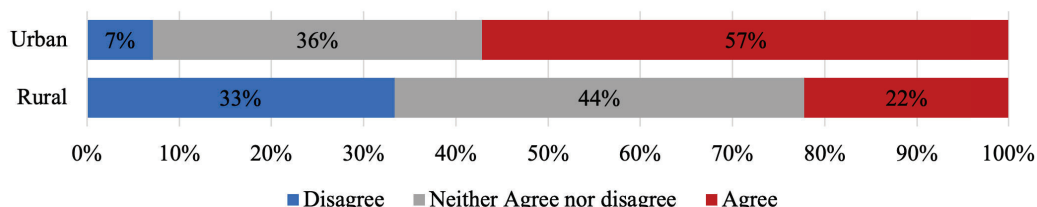


Note: The data reveal judges' level of trust in risk assessment tools in urban and rural counties. *Source:* Judge Survey

levels may well influence the extent to which they trust such interventions or make use of them in their decision-making.

Finally, judges from urban counties were more likely than rural judges to indicate that ARAs allow them the freedom to make decisions that may be unpopular (see Figure 18).³¹ This may be explained by the political composition of the constituencies in urban and rural areas. Urban and large counties in Ohio tend to have a population with more diverse political views, while rural areas tend to have a more homogeneous political composition.³² Judges in these areas need to respond to a more diverse population, requiring more objective information to justify decisions that may be unpopular to some and popular to others.

Figure 18. ARAs allow me freedom to make unpopular decision by urban and rural counties



Note: The data indicate judges' opinions on whether risk assessment tools provide them with the freedom to make unpopular decisions in urban and rural counties. *Source:* Judge Survey

These results suggest there are differences between how judges from relatively larger or smaller

³¹Mann-Whitney test ($z = -2.195, \rho = 0.03$).

³²We used the results of the governor election from November 2022. For each county we estimated how big the gap was between the first and second majorities. Counties in which the difference is bigger have a politically more homogeneous population. For testing this measure against the statement regarding ARAs allowing judges freedom to make unpopular decisions, we used a Mann-Whitney test ($z = -0.32, \rho = 0.05$). We also tested the relationship between the created measure of political difference and the variables for urban and large counties using a Pearson correlation for urban areas (-.67) and for large counties (-0.62), both statistically significant at 95%.

counties view and use risk assessment tools. While there could be several normative and environmental factors that contribute to such distinctions in views and frequency of use, which are beyond the scope of this project, it is worth noting that such differences will inherently create variability and inconsistency in how risk assessment tools are used in practice across jurisdictions within the state of Ohio.

4. Best Practice, Ohio Practice, and Recommendations

In this section, we delve into the analysis of best practices concerning the use of ARA scores in bail and sentencing decisions. Drawing from the literature, we will outline the identified best practices and contrast them with the observed practices in Ohio. By examining these practices, we aim to discern whether Ohio aligns with established best practices in this domain and so generate recommendations as to how Ohio could better align its use of ARAs with existing ideas on best practices.

4.1 Interpreting and Communicating Risk Assessment Tool Results

4.1.1 Best practice and Ohio practice

Presentation of Risk Information. To facilitate easier interpretation of risk information, the literature recommends presenting risk information as percentage estimates with confidence intervals and framing it in terms of the probability of not recidivating (positive framing) to improve interpretation (Hamilton, 2020). However, the ORAS tools do not follow these best practices. Instead, they present risk as a numerical score (0-9 for ORAS-PAT and 0-49 for ORAS-CST), which is then categorized into low, moderate, high, and (for CST) very high risk groups. Judges typically receive these categorical risk groups, though interviews revealed inconsistency in whether numerical scores are always provided. Categorical scores are more absolute and don't convey a defendant's relative risk compared to others.

Additionally, during interviews, judges displayed varying levels of understanding regarding the ARA results in the PSI for sentencing. Some judges could clearly distinguish between sections directly tied to the ARA, supplementary information, and more detailed explanations, while others

found these boundaries less clear. One judge explained how they used the ARA score:

I think I'll just go to the PSIs. That's the ORAS. So we use the ORAS in our PSI, I think it's a huge benefit for us as judges to review what ORAS has showed to help me to, I'm comfortable with saying "Hey! you're going to be in group C supervision and follow the terms and conditions of the probation officer," as opposed to saying, do this or go here or take these classes or go to this drug rehab because ORAS has set up what it looks like to be a group C reporting requirement (Interviewee No. 1).

Risk Assessment Tools are Not to be Determinative. The literature recommends that ARAs should serve as a tool to inform decision-making, rather than being the sole or automatic determinant of a decision (Hamilton, 2020; Latessa et al., 2018). Our interviews with judges revealed a consistent understanding that ARAs are one piece of information considered alongside other factors in decision-making. Several quotes from judicial interviewees support this perspective:

[T]he score is just one other factor to consider. And, generally speaking, I think that's the way that it should be. I mean you shouldn't just say...automatically this is the type of sentence that should be imposed, or this is the type of bail that should be imposed (Interviewee No. 14).

You can't blindly rely on these scores. They're just tools... Data points can be wrong, errors can happen, and so no one should blindly rely on a score when making any sort of decision involving human beings (Interviewee No. 13).

I don't take the risk of assessment score as gospel... it's a number that... it's one of many pieces of information that I'm looking at (Interviewee No. 3).

I look at all the other factors such as prior history and history of court appearance. But like I said, I think the tool kind of distills all that in a way that I can take it as almost as a, I don't want to say like a cheat sheet or a cheat code, but unless there's

something out there that if the offense involves a minor victim, that may not necessarily be reflected in the tool. So, I can account for that. But otherwise, if we're talking about drug possession offense, trafficking offense, or property offense, I'm probably nine times out of ten going to go along with the recommendation based on the tool, based on the risk assessment tool (Interviewee No. 9).

A common concern in sentencing, echoed by stakeholders interviewed in this study, is the principle that individuals should be sentenced based on their specific actions and circumstances, rather than on group or aggregated factors. Representatives from victims' rights and civil liberties organizations manifested concern with the tools' use of aggregated statistics to judge the risk of an individual, which might obscure unique individual factors which are important:

I think the tool is useful. But you know, like I said, it should be a factor in the decision, not the factor (Interviewee No. 9).

Sometimes there may be people who pose a high risk but can still be released if we put certain things around them to help them succeed. And I worry that that gets lost on some judges if we start moving to these analytical tools that you know, purport to do more than they really can do. You know, judges still have to judge, and part of judging is seeing the whole person, not just a damn number and not just the damn algorithm that says this person is going to be a good bet and this person's not (Interviewee No. 10).

Judges' Interpretation and Understanding of Risk Scores. The literature also notes that risk information should be communicated in a way that is “interpretable to the decision makers who receive them” (Hamilton, 2020, p. 52), and that such information should be presented in a “clear, comprehensible manner” (Garrett and Monahan, 2020, p. 488). Given the presentation of categorical risk scores provided by the ORAS tool, it may be possible to conflate such categorical risk scores with a relative percentage (e.g., to interpret a “moderate” score as approximately 50% likelihood of recidivism). In order to investigate judges' understanding and knowledge of how

categorical risk scores translate into recidivism risk, we asked judges to select, from a list of multiple options, what percentage of risk for recidivism is indicated by a “High Risk” score for a defendant using the ORAS-PAT tool, and “Very High Risk” for a male defendant using the ORAS-CST tool (these are the highest risk categories for each tool; the ORAS-CST tool has separate risk rankings for male and female defendants) (see [Latessa et al., 2009](#)). In response, we noted that only one respondent selected the correct corresponding recidivism risk percentage for the ORAS-CST tool, and no judge correctly guessed the corresponding recidivism risk percentage for the ORAS-PAT tool (see Table 2).

Table 2. Accuracy of judges’ responses to how categorical risk scores equate to percent risk of recidivism, according to initial ORAS tool validation

| | Did not answer | Did not know | Responded correctly | Overestimated the risk | Underestimated the risk |
|--|-----------------------|---------------------|----------------------------|-------------------------------|--------------------------------|
| “High Risk”: ORAS-PAT Tool | 46.3% | 34.1% | 0.0% | 15.0% | 5.0% |
| “Very High Risk”: ORAS-CST Tool | 26.8% | 56% | 2.4% | 12.1% | 2.4% |

Note: The data shows the percentage of judges who correctly translated categorical risk scores into the corresponding percent risk of recidivism. *Source:* Judge Survey

It is important to note that many respondents either didn’t know the correct answer or left these questions unanswered, potentially indicating a lack of knowledge. Those who did respond incorrectly tended to overestimate the recidivism risk associated with each categorical ranking. This suggests that judges may not fully understand the absolute risk these categorical scores represent and could be overestimating the actual recidivism risk posed by defendants in these categories.

We do want to caution that these results are only from a portion of the overall judge population, and that we only asked judges about the highest risk categories. Thus, responses could be more accurate for other categories. We also recognize that it is difficult to expect judges to have perfect memory for such statistics. However, the fact that we provided judges with an even distribution of options for both questions (e.g., options for ORAS-CST question were: 100%, 88%, 69%, 50%, 34%, 16%, “I don’t know”),³³ and yet judges were almost uniformly unable to identify the correct

corresponding level of risk for recidivism, is striking.

These results clearly indicate there is a gap in judges' knowledge regarding the level of risk represented by each category. Given the small number of responses, we cannot determine with certainty whether judges would tend to over or underestimate the risk percentage. In the same way, we cannot rule out the impact that this misperceived risk can have on judges' decision-making, and ultimately on defendants' future.

4.12 Recommendations: Communicating Results

- Consistent with best practices (see [Hamilton, 2020](#); [Latessa et al., 2018](#)), the judges in our data set treated ARA scores as just one of the many pieces of information to consider in making bail or sentencing determinations. Nevertheless, courts could still consider reinforcing this positive behavior by adding a clear and visible statement on PSI reports indicating that risk assessments scores are but one of a variety of factors the judge should consider.
- Interpretation of risk assessment scores can be challenging for judges. We recommend developing a standard for presenting risk assessment results in a clear and comprehensible manner. Additional research is necessary to establish the most effective methods for communicating risk assessment scores. This may involve presenting total scores as numbers or percentages, segmenting scores by section, assigning individuals to different risk categories, indicating the percentage of risk (recidivism or other) associated with specific scores, including 95% confidence intervals, offering an overview of the risk distribution in general, or some combination of these approaches (see [Hamilton, 2020](#)).

4.2 Training in the Use of ARAs

A key best practice for applying risk assessment tools is ensuring that those administering the assessments, including judges and other criminal justice professionals, are properly trained. Even the most well-designed and validated tools will only be effective if they are implemented and used

³³Correct recidivism risk percentage for ORAS-PAT score of “High Risk” = 30%; correct recidivism risk percentage for ORAS-CST score of “Very High Risk” = 69% (see [Latessa et al., 2009](#)).

correctly in practice (Stevenson, 2018). Training plays a crucial role in how ARAs are applied in local courts. Several best practices have been proposed to guide what effective training for using ARAs should entail.

4.21 Best practice and Ohio practice

Training for those administering risk assessment tools. Best practices have called for those who use the tools and develop the scores to receive training in how to do so (Bavitz et al., 2018; Latessa et al., 2018). Our survey of court staff members revealed that probation officers, who are primarily responsible for administering tools in most jurisdictions,³⁴ and, where appropriate, other specialized staff whose primary duty involves administering ARA tools (such as pre-trial services and PSI writers), reported receiving adequate training in all but one of the courts that responded (the respondent for the other court said that they did not know). The literature suggests that training should be provided to all users of the ARA tools, including judges, attorneys, pretrial services, clerks, and other relevant stakeholders. (see DeMichele et al., 2021; Hamilton, 2020). In terms of content, the training should include information to help those users better understand the “nature and limitations of the tools” (Partnership on AI, 2019, p. 26). The ORAS tool includes interviews as part of the assessments, then it is also recommended that those scoring defendants on the tool also receive “sufficient training in how to reduce interviewer bias and on culturally sensitive interview skills” (Hamilton, 2020, p. 39). Discussions with ODRC staff indicated that these procedures were indeed being followed as suggested by best practices:

[T]raining for that interview person...there is a whole section on effective interviewing as part of the training. And we talk about bias we talk about, you know, if you have a hostile client or somebody who doesn't want to talk, different techniques. But then we talk about not asking biased questions. . . as part of that training there is a whole section on effective interviewing and how to eliminate some of those, you know, interview bias or things that will shut down an interview (Interviewee No. 1).

³⁴Data indicate that about 56% of probation officers are trained frequently.

Regarding ORAS specifically, the original designers of the tools at the University of Cincinnati created tailored training materials and recommended that “ODRC follow the protocol developed by the University of Cincinnati for training personnel on the assessment instruments” (Latessa et al., 2009, p. 46). Best practices have also called for “retrainings at reasonable intervals...for evaluators to maintain skills” (Hamilton, 2020, p. 39). These practices appear to be in place and actively followed for those using the ORAS tool in Ohio courts. ODRC officials stated the following.

Regarding original training:

We go through the manuals and the scoring guides in depth, and then you would see that a lot of the training is, you’re actually watching videos. You score [the video interview] out independently, you score it in a group, and then we come together, and we talk about each one of the items that score out and we tell you what the correct answer is and why, and give them all, you know, the opportunity to ask any questions to clarify. And then, at the end of all that, you actually take a video test based on an interview and score it out (Interviewee No. 1).

The [ORAS] certification is valid for 3 years...[I]t’s a 2-day training and if for some reason someone doesn’t pass, there are other options after that, that are shorter, but more...in a smaller class size to give more specific attention to those individuals.” (Interviewee No. 17).

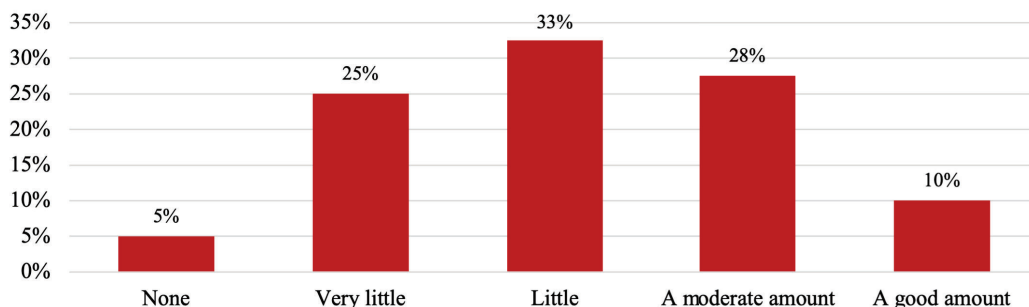
Regarding recertification:

[A]t that 3-year mark, there is a recertification process...there’s 2 attempts online; if those individuals can’t pass after the second attempt, then we do meet in person with them to recertify them (Interviewee No. 17).

Training for Judges. Judges are the ones who decide whether, and how, to account for risk assessment information in their decisions regarding defendants. Accordingly, best practices have called for judges to receive training on how to interpret and properly use ARA tools (Bavitz et al.,

2018). Our results indicated that such training was not uniformly provided to judges. As can be seen in Figure 19, the majority of judges reported receiving little to no training, or a moderate amount at best.

Figure 19. Amount of training judges reported receiving regarding ARA tools



Note: The data reveal the reported amount of training judges receive on risk assessment tools. *Source:* Judge Survey

Interviews and survey responses revealed that training on ARAs was typically provided by external organizations, such as the Ohio Judicial Conference, rather than the court or ODRC. This training was irregular and more common when the ORAS tool was first introduced. Notably, there was a significant positive relationship between judges' years of service and the amount of training they reported receiving.³⁵ This suggests that newer judges have received less training. While this may be partly due to fewer training opportunities for judges with less experience, it could also be a result of the extensive training offered when the ORAS tools were first introduced, which was likely more accessible to longer-serving judges at that time. One interviewee supported this idea by stating:

[B]ack in 2008, when we were trying to move the court to look at the tool of risk assessment as it related to probation, there was training...but I don't think it's being done now (Interviewee No. 21).

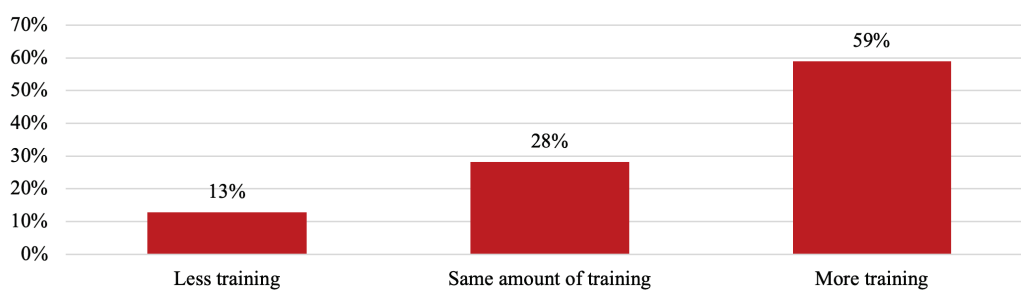
When asked about their preference for training on risk assessment tools, most judges expressed interest in receiving more (see Figure 20). This indicates that judges are generally receptive to further training, a sentiment echoed by several judges in our interviews:

³⁵Relationship tested using Spearman correlation ($Rho = 0.33, \rho = 0.0405$).

We need training...I think training should be standardized in the state, for whatever it is. . . the tool we're using, or why we're doing it (Interviewee No. 21).

Oh, yeah, no, absolutely [I'd be open to more training] to help us all understand. We've had, you know, trainings on bail and bond, obviously through the judicial college and have had discussions as to risk assessments as a tool to use in making bond decisions, but we've not had training on, you know kind of here's how we arrived at the ORAS scores (Interviewee No. 14).

Figure 20. Amount of future training judges reported being interested in receiving



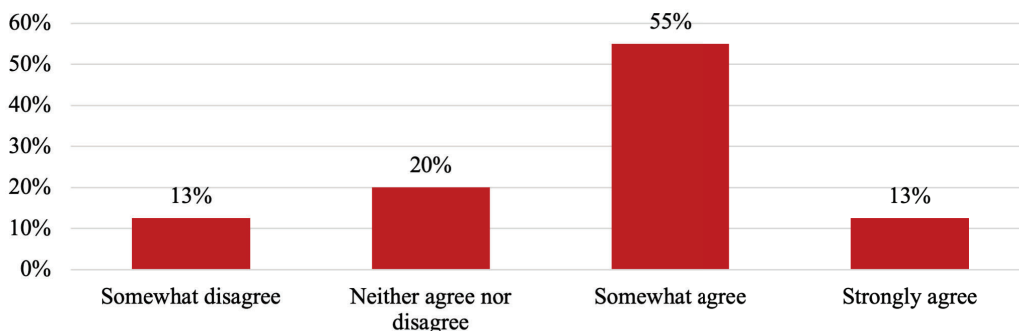
Note: The data reveal the reported amount of future training judges would like to receive on risk assessment tools.
Source: Judge Survey

This openness to training reduces as age and experience increase, and as judges' perceptions of their own ability to assess risk, compared to ARAs, increases. We found all these relationships to be negative and significant. These results are not surprising, as one could expect that younger and less experienced judges, having likely received less training, would report a higher need for it. In the same way, if judges perceive their own judgment as superior to an ARA, they will not only give these tools less importance and use them less frequently, as stated previously, but will also likely report needing less training.

We asked judges the extent to which they felt "well-equipped to understand what risk assessment scores communicate about defendants" Overall, judges mostly agreed with this statement, indicating that they feel well-equipped to understand ARA scores (Figure 21). These results contrast with the judges' reported desire for future training. We see that 59% of respondents report a need for more training, while 68% feel to some extent equipped to understand the ARA scores.

These results could indicate that the judges' declared need for future training may be overestimated due to bias, as being open to learning can be thought of as a socially desirable attitude. This means judges could tend to overreport their desire for future training, when in reality they do feel competent to use ARAs.

Figure 21. Extent to which judges agree with the statement: "I feel well-equipped to understand what ARA scores communicate about defendants."



Note: The data reveal judges' confidence in their understanding of risk assessment scores. *Source:* Judge Survey

We also saw that there was a significant positive correlation between the amount of training judges received and this measure of how well equipped they feel to understand ARA scores.³⁶ Similarly, we see a positive association between years of experience serving and the extent to which they feel equipped to understand ARAs.³⁷ This provides some evidence that training and experience are effective in terms of allowing judges to feel better prepared to properly use ARA scores to judge defendants' risk.

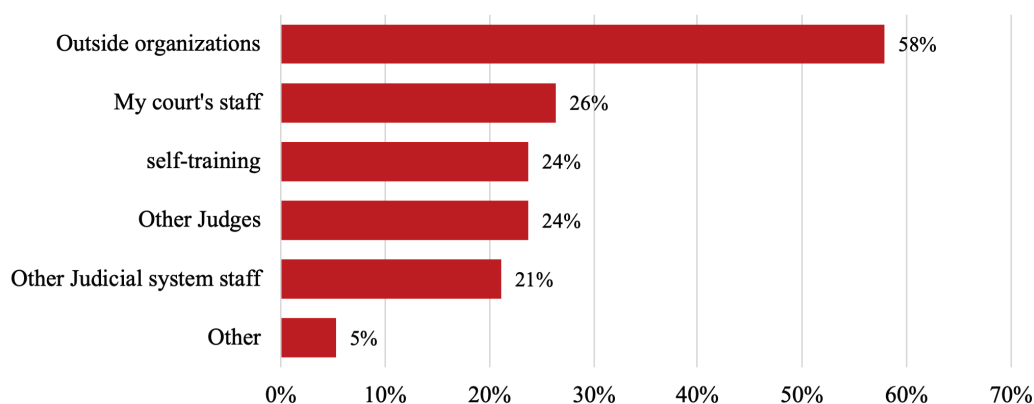
Figure 22 below shows that most respondents receive training from outside organizations. Almost 60% claim that they received risk assessment training from outside organizations. Approximately 25% of respondents mentioned receiving training from court staff, themselves, other judges, and other personnel within the judicial system. Only 5% of respondents receive training from other providers.

Our analysis also shows that the distinction of who provided the training was only marginally connected to the perceptions of how well-equipped judges feel, as well as their desire for future training. Specifically, judges were more likely to feel well-equipped to understand what risk as-

³⁶Relationship tested using Spearman correlation ($Rho = 0.46, \rho < 0.01$).

³⁷Relationship tested using Spearman correlation ($Rho = .52, \rho = 0.00$).

Figure 22. Who provided the training



Note: The data show the sources of training on risk assessments for judges, broken down by percentage. *Source:* Judge Survey

assessment scores communicate when trained by other judicial system staff.³⁸ Similarly, judges that have been trained by outside organizations, such as ODRC or the Ohio Judicial Conference, also have a lower desire for future training.³⁹

It is also worth noting that the judges' desire to receive more training was correlated with several measures of attitudes towards the tools.

-How much training they have received and 1) frequency of use for decision-making: $Rho = 0.26$.

-Desire for future training and 1) ARAs enable me to better judge defendants' risk: $Rho = 0.38$, 2) ARAs help guard own bias: $Rho = 0.45$, 3) ARAs allow me to be more consistent in my decision-making: $Rho = 0.37$, 4) ARAs importance for sentencing: $rho = 0.38$.

-How well equipped they feel and 1) ARAs enable me to better judge defendants' risk: $Rho = 0.29$, 2) ARAs add something to my decision-making: $Rho = 0.37$, 3) ARAs importance for sentencing $Rho = 0.44$. Specifically, judges who believed the tools enable them to better judge risk posed by defendants, help them guard against their own biases during decision-making, and believe the scores help them to be more consistent in their decision-making, were more likely to say they would like to receive more training in the future. This indicates that judges who find the tools to be helpful and useful for decision-making, and thus see value in their use, are the ones who tend to desire more training. Conversely, those who feel the tools are less useful or who

³⁸The relationship was explored using the Mann-Whitney test ($z = -1.85, \rho = 0.06$).

³⁹The relationship was explored using the Mann-Whitney test ($z = -1.89, \rho = 0.07$).

don't rely on them as much, were less likely to express a desire for more training. In the same line of analysis, we found that neither opinion, frequency of use, nor importance for both bail and sentencing are significantly correlated with the amount of training judges have already received. These findings suggest that judges' reasons for not using ARAs or seeing them as helpful are not primarily because they do not feel they have enough training, but rather are likely related to other factors.

4.22 Recommendations: Training of Judges and Assessors

Training of assessors:

- Those who administer ARA tools receive adequate training at regular intervals. Courts should continue providing mandatory training and recertification for administrators. Regular retraining sessions are crucial for enabling evaluators to maintain their skills.
- Courts should standardize training protocols for all personnel administering risk assessment tools. In addition to instructing evaluators on how to use the tools, the training should include information on the nature and limitations of risk assessments, effective interviewing techniques to mitigate bias, and cultural sensitivity. Training should be updated regularly to incorporate new issues, research, and technical updates.
- Court systems should allocate sufficient resources for training evaluators.

Training of judges:

- Our findings suggest that judges do not receive adequate training on ARA tools and scores. Sixty percent of judges would like to receive more such training. We accordingly recommend the provision of training programs specifically designed to educate judges.
- Such training programs should include information on how risk assessment tools arrive at risk scores and on the uses for which a particular tool is, and is not, intended. Based on the data presented in the Communication section on judges' difficulty in matching risk categories with recidivism risk percentages, training programs should include clear information on this

topic. Additionally, training programs should summarize the literature regarding potential bias in ARA tools and how this may affect risk scores for different populations. Finally, training should include information on the ways in which algorithmic risk assessments can undermine Due Process values and on the steps that judges can take to ameliorate this.

- The Ohio Judicial Conference has been a forum for ARA training in the past and could play this role again.
- Court systems should recognize that judges newer to the bench or who serve in rural areas have likely had fewer opportunities for training, and should seek to address these disparities and ensure that judges have access to adequate training. To accommodate different learning preferences and logistical constraints, courts should consider offering training opportunities in various formats such as in-person, online, or hybrid. Mandatory training for judges, such as that required for evaluators, should be considered.

4.3 Ensuring Due Process

4.31 Best practice and Ohio practice

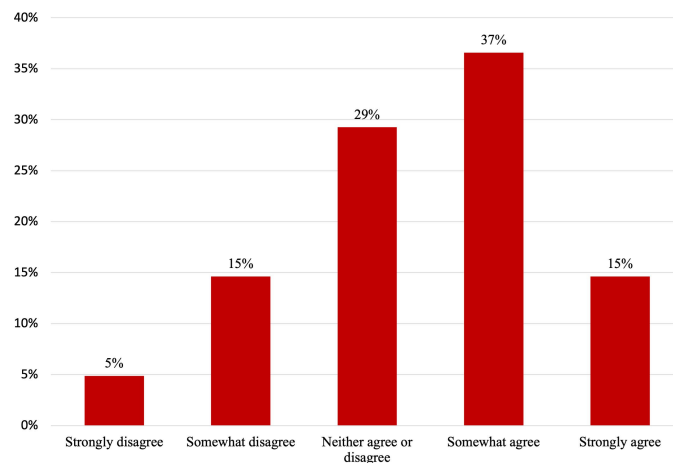
The government can deprive a person of property, liberty and, in some circumstances, even life. But under the Fifth and Fourteenth Amendments of the U.S. Constitution, it can only do so if it acts in accordance with due process of law. In the criminal context, this generally means that the government must provide the defendant with notice of the proposed action and the grounds for it, an opportunity to challenge those grounds, and a decision by a neutral decision-maker (Kuckes, 2006). Courts provide these protections more fully during criminal trials than in pre-trial proceedings (Id.). Still, the ideas of notice and an opportunity to be heard animate ideas of fair procedure throughout the criminal prosecution process.

In the ARA context, due process ideals would suggest that defendants (and prosecutors) be provided with notice of the risk score and how the court intends to use it; enough information about that score that they can meaningfully challenge it; and an opportunity to make that challenge (Eckhouse et al., 2019; Hamilton, 2020; Partnership on AI, 2019). Emerging case law provides

some support for these best practices. At least one State Supreme Court has held that due process requires the state to provide a defendant with the specific risk assessment questions and answers and the scores associated with them ([Kansas v. Walls, 2017](#)). However, another State Supreme Court has held that the defendant is not entitled to the proprietary algorithm, developed by a private company, that the state uses to produce the ultimate risk assessment score ([State v. Loomis, 2016](#)).

In assessing how Ohio practices compare with due process best practices, we began by asking judges if they agree with the statement, “risk assessment tools pose due process concerns.” Our survey results largely indicated that judges view ARA tools as presenting significant due process concerns (see Figure 23), although there was some ambiguity, **as some judges neither disagree nor agree**. Our interviews with judges served to confirm this general sentiment, while also revealing that many judges indicated they had not thought extensively about due process concerns related to ARAs. Thus, the ambiguity seen in our survey results may reflect that judges did not feel they had thought adequately about the topic to say one way or another.

Figure 23. The extent to which judges agree with the statement: “ARA tools pose due process concerns.”



Note: The data reveal judges’ opinions on whether risk assessment tools pose due process concerns. *Source:* Judge Survey

Access to Information and Challenging ARA scores. For a party to be able to challenge the information on which a given risk score is based, that information must continue to exist. Best practices thus indicate that this data be retained for the period during which such a challenge might be brought (see [Hamilton, 2020](#)). Ohio appears to satisfy this best practice. In their interviews,

ODRC staff made clear that risk assessment information almost always went through a central online system, and was thus retained by ODRC, though there were a few exceptions to this:

[E]verybody's required to enter it in [the computer system]. There are some exceptions to that...small counties where they're doing it, old school paper form, and they may not be entering it in the system, but by and large most all programs are entering it in the system (Interviewee No. 1).

To comport with due process values, defendants should have the access to, and the opportunity to challenge, the ARA scores themselves and the information used in the assessment. The judges who responded to the survey reported a wide variety of practices with respect to who receives this information, when they received it, and what information they receive (see Table 3). This data shows relatively broad access. But it also suggests a lack of consistency with respect to judicial practice in this crucial area.

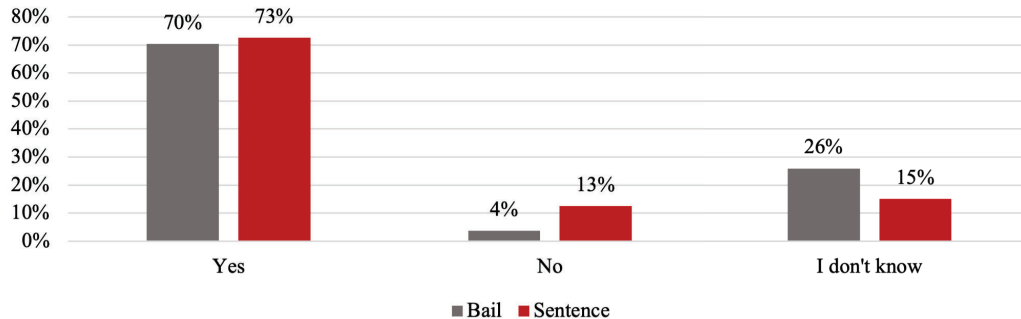
Table 3. Who receive ARA, when, and what information do they receive

| Party | Receive ARA | When | | | | | What information receive | |
|----------------------------|-------------|----------------------|---------------------------------------|-------------------|-------------------------|-------|--------------------------|-------------------------------------|
| | | Prior to arraignment | At the start of or during arraignment | After arraignment | Only if they request it | Other | Just the score | Score and information underlying it |
| Bail | | | | | | | | |
| Defense Counsel | 56% | 53% | 27% | 0% | 7% | 13% | 27% | 73% |
| Prosecution | 56% | 57% | 21% | 0% | 7% | 14% | 21% | 79% |
| Caseworker, Parole Officer | 19% | 75% | 0% | 0% | 0% | 25% | 25% | 75% |
| Other | 15% | 25% | 0% | 25% | 0% | 50% | 0% | 100% |
| No One Else | 19% | | | | | | | |
| I don't know | 11% | | | | | | | |
| Sentencing | | | | | | | | |
| Defense Counsel | 75% | 20% | 7% | 20% | 0% | 53% | 37% | 63% |
| Prosecution | 73% | 21% | 3% | 14% | 3% | 59% | 38% | 62% |
| Caseworker, Parole Officer | 30% | 8% | 8% | 17% | 8% | 58% | 8% | 92% |
| Other | 8% | 0% | 0% | 0% | 0% | 100% | 50% | 50% |
| No One Else | 13% | | | | | | | |
| I don't know | 8% | | | | | | | |

Note: The data illustrates which parties receive risk assessment scores, when they receive them, and what information is included. *Source*: Judge Survey

As for the ability to challenge ARA scores during legal proceedings, 73% of judges indicated ARA scores could be challenged at the bail decision point, and 70% of judges indicated they could be challenged at sentencing (see Figure 24). However, when asked in more depth who can challenge the scores, their responses again showed great variation, ranging from “Everyone” to only one specific party.

Figure 24. Can ARA scores be challenged?



Note: The data show judges' views on whether risk assessment scores can be challenged or not. *Source:* Judge Survey

The interviewees confirmed this ability to challenge, but made clear that such challenges are extremely rare, perhaps due to defendants' lack of knowledge about ARA scores and how to analyze and challenge them.

It is challenged occasionally, but not very often. But to answer your question, yes, it could be challenged. It's not done often (Interviewee No. 21).

I've never seen an ORAS score be challenged (Interviewee No. 13).

They have the opportunity to challenge it. But most don't have enough knowledge to challenge the risk assessment. What they would challenge probably would be more specifics in, you know, well, this has changed since the score was done (Interviewee No. 12).

4.32 Recommendations: Aligning ARAs with Due Process Values

- The survey responses suggest that judges are not yet sufficiently aware of due process concerns related to ARAs. Judges should accordingly be provided information on the ways in which algorithmic risk assessments can undermine due process values and on the steps that judges can take to ameliorate this.
- The state should identify the period of time during which a defendant might want to challenge their risk assessment score and should recommend to the counties that, after conducting a risk assessment, they retain the underlying information for that period.
- There should be more consistency with respect to when, how, and to whom information about ARA scores is provided. A central body should develop a standard practice, share it with each of the Ohio counties, and encourage the counties to adopt it.

4.4 Oversight

4.41 Best practice and Ohio practice

An important aspect of how algorithmic risk assessment tools are used in practice entails processes for feedback and quality management. Such processes ensure that ARAs are being used appropriately in practice and allow for corrections in instances where this is not the case. [Garrett and Monahan \(2020\)](#) suggested that such monitoring of practical implementation is key, and that “providing feedback to judges concerning their use of risk assessment [sic] may improve their performance” (p 480).

The first area of feedback that we encountered was in the process of bi-annual audits conducted by ODRC for each Ohio County’s use of risk assessment tools. As ODRC staff noted in interviews, the process is every other year, and then includes a process for suggestions based on the audit in the intervening years:

We audit one year and then, the subsequent year, we do a coaching process to help you bring any non-compliant standards into compliance (Interviewee No. 1).

Along with this, one of the primary aspects that ODRC focuses on during this audit and coaching process is ensuring some form of quality assurance. This generally includes things such as double-coding and assessing inter-rater reliability.

[W]e have standards that require some sort of quality assurance process on those conducting, or, as and other assessments is, is that does that include things, you know, such as like double coding inter-rater reliability, those types of things, the process you would hear us recommend, and programs don't have to do. It is a direct, direct observation of the interview, and then a double coding (Interviewee No. 1).

Best practices recommend such double-coding and checking inter-rater reliability at regular intervals (Hamilton, 2020). Thus, it seems that Ohio is, for the most part, adhering to this standard. There are times, however, when counties with smaller staff are not able to complete such quality assurance procedures, especially double coding. In such cases, ODRC may encourage counties to work together:

If it's a one-person program, we're probably [going to] encourage you to collaborate with your neighbor, who is also a one-person program, and you do quality assurance for each other...your neighboring county, that kind of thing. So that's kind of a typical recommendation when we see those kinds of things (Interviewee No. 1).

One other important aspect of oversight for the use of ARA tools in practice involves the use of discretionary overrides. Many risk assessment tools, including ORAS, allow for those conducting the assessment to manually override a risk score. Best practices have called for such overrides to be subject to additional oversight, and for detailed explanations to be provided in such cases (Chohlas-Wood, 2020; Hamilton, 2020). In the case of ORAS, such discretionary overrides are limited to 10% of cases, and require that an explanation for the override be provided as part of the information entered into the computer system:

[Y]ou can override, but you would hear us always say, to base it off of a secondary assessment, not just some kind of gut feeling. And the other thing we say is that

generally you should not be overriding more than about 10% of the people that you see (Interviewee No. 1).

ODRC also indicated that while a county may go over the 10% discretionary override threshold in occasional instances, they are usually below the threshold in the subsequent quarter that overrides are tracked. ODRC does have the ability to levy financial sanctions on courts that have a consistent problem with going beyond the override cap, but they have never had to use such sanctions.

One way that differences in court size may manifest itself is in how use of the tools is monitored. Quality management processes such as double-coding and assessing inter-rater reliability may be difficult or impossible in jurisdictions where limited court staff are eligible for ORAS training and administration. While such processes can be accomplished through coordination with other nearby jurisdictions of similar size, this adds additional burdens to the completion of such processes, as they now involve travel and aligning schedules of multiple jurisdictions.

[T]here's a lot of places that if you do the [Pre-Sentence Investigation Report, which would include the ORAS], you're probably going to get the offender, too. And you know the ORAS, in most jurisdictions, will generate what your contact standards are on the other end; how often you have to do more work. And so, if I'm doing the assessment, it's probably self-serving in a lot of cases for me to underscore the offender so that I don't have as much work to do if they get probation or over score the offender to make them look worse, and maybe the judge sends them to prison, so now I don't get the case. So yeah, I think who's doing the risk assessment can yield different results, especially depending on whoever's doing it and what their mindset is. If you are in a department, and you feel as though you're overworked already and undervalued, you're probably more likely to not, you know, expend the effort to do the ORAS the way it's supposed to be done (Interviewee No. 20).

This raises concerns that such quality assurance processes may vary across jurisdictions based on the size of the court staff. Relatedly, it also seems that variations in resources (both staffing and

financial) may create differences in how risk assessment tools are administered. For example, larger counties may have dedicated PSI writers, or individuals who specifically conduct risk assessments. However, in small counties, all these duties might fall to just a few individuals in the probation department. In addition to the problems a heavy caseload can have in the quality assessment, the lack of dedicated PSI resources has the potential of creating conflicts of interest. If the individuals conducting the assessment belong to the same unit or department responsible for supervising the defendant, they may lack impartiality, as the outcome of the assessment could impact their future workload.

4.42 Recommendations: Oversight and Accountability

- Ideally, an independent third party would evaluate ARA scoring systems for sufficient accuracy. Where resources or other constraints make this infeasible, court systems should enhance accountability through double-coding, or random/risk-based multiple raters testing, and should ensure that sufficient resources are available for these measures.
- Court systems should, wherever possible, assign the risk assessment of a given defendant to someone who will not be assigned to supervise that same defendant if released. Otherwise, evaluators will have incentives to score a defendant more or less highly based on their or their team's current workload.
- Court systems should review the impact of external incentives (i.e., 10% discretion caps) to ensure that they avoid a vicious cycle where excessive overrides results in fewer resources, which leads to lower quality risk assessments, which in turn produces more overrides.
- Court systems should consider incentivizing oversight practices, training for judges, and other practices that improve the implementation of ARAs.
- Court systems should further adopt the practice, already pursued by some courts, of asking judges and courts administrators for their views on ARAs and then using that information to fine tune training and ARA implementation.

5. Recommendations

This report has provided a number of recommendations. Here, we pull these recommendations together and list them as a complete set:

5.1 Communicating Results

- Consistent with best practices, the judges in our data set treated ARA scores as just one of the many pieces of information to consider in making bail or sentencing determinations. Courts should consider reinforcing this positive behavior by adding to the presentation of ARA results a statement that the risks scores are but one of a variety of factors the judge should consider.
- Interpretation of risk assessment scores can be challenging for judges. We recommend developing a standard for presenting risk assessment results in a clear and comprehensible manner. Additional research is necessary to establish the most effective methods for communicating risk assessment scores. This may involve presenting total scores as numbers or percentages, segmenting scores by section, assigning individuals to different risk categories, indicating the percentage of risk (recidivism or other) associated with specific scores, including 95% confidence intervals, offering an overview of the risk distribution in general, or some combination of these approaches.

5.2 Training of Assessors and Judges

5.21 Training of assessors

- Those who administer ARA tools receive adequate training at regular intervals. Courts should continue providing mandatory training and recertification for administrators. Regular retraining sessions are crucial for enabling evaluators to maintain their skills.
- Courts should standardize training protocols for all personnel administering risk assessment tools. In addition to instructing evaluators on how to use the tools, the training should

include information on the nature and limitations of risk assessments, effective interviewing techniques to mitigate bias, and cultural sensitivity. Training should be updated regularly to incorporate new issues, research, and technical updates.

- Court systems should allocate sufficient resources for training evaluators.

5.22 Training of judges

- Our findings suggest that judges do not receive adequate training on ARA tools and scores. 60% of judges would like to receive more such training. We accordingly recommend the provision of training programs specifically designed to educate judges.
- Such training programs should include information on how risk assessment tools arrive at risk scores and on the uses for which a particular tool is, and is not, intended. Based on the data presented in the Communication section on judges' difficulty in matching risk categories with recidivism risk percentages, training programs should include clear information on this topic. Additionally, training programs should summarize the literature regarding potential bias in ARA tools and how this may affect risk scores for different populations. Finally, training should include information on the ways in which algorithmic risk assessments can undermine due process values and on the steps that judges can take to ameliorate this.
- The Ohio Judicial Conference has been a forum for ARA training in the past and could play this role again.
- Court systems should recognize that judges newer to the bench or who serve in rural areas have likely had fewer opportunities for training, and should seek to address these disparities and ensure that judges have access to adequate training. To accommodate different learning preferences and logistical constraints, courts should consider offering training opportunities in various formats such as in-person, online, or hybrid. Mandatory training for judges, similar to that required for evaluators, should be considered.

5.3 Aligning ARAs with due process values

- The survey responses suggest that judges are not yet sufficiently aware of due process concerns related to ARAs. Judges should accordingly be provided information on the ways in which algorithmic risk assessments can undermine due process values and on the steps that judges can take to ameliorate this.
- The state should identify the period of time during which a defendant might want to challenge their risk assessment score and should recommend to the counties that, after conducting a risk assessment, they retain the underlying information for that period.
- There should be more consistency with respect to when, how, and to whom information about ARA scores is provided. A central body should develop a standard practice, share it with each of the Ohio counties, and encourage the counties to adopt it.
- Court systems should provide defendants and other stakeholders with enough education about algorithmic risk assessments that they feel capable of understanding and, where necessary, challenging ARA scores.

5.4 Oversight and accountability

- Ideally, an independent third party should evaluate ARA scoring systems for sufficient accuracy. Where resources or other constraints make this infeasible, court systems should enhance accountability through double-coding, or random/risk-based multiple raters testing, and should ensure that sufficient resources are available for these measures.
- Court systems should, wherever possible, assign the risk assessment of a given defendant to someone who will not be assigned to supervise that same defendant if released. Otherwise, evaluators will have incentives to score a defendant more or less highly based on their or their team's current workload.
- Court systems should review the impact of external incentives (i.e., 10% discretion caps) to ensure that they avoid a vicious cycle where excessive overrides results in fewer resources,

which leads to lower quality risk assessments, which in turn produces more overrides.

- Court systems should consider incentivizing oversight practices, training for judges, and other practices that improve the implementation of ARAs.
- Court systems should further adopt the practice, already pursued by some courts, of asking judges and courts administrators for their views on ARAs and then using that information to fine tune training and ARA implementation.

6. Conclusion

Ohio, like many other states across the nation, makes extensive use of algorithmic risk assessment tools in its bail and sentencing proceedings, particularly in the Ohio Court of Common Pleas. This report has conveyed the results of surveys of Ohio Common Pleas judges, court administrators, and probation officers, as well as interviews of judges and a diverse array of stakeholders. It has described how Ohio judges view algorithmic risk assessment scores, and how they and the others implement and use them. At each stage of the analysis, the report has compared Ohio practice to best practices identified in the literature. The results are revealing. They show important ways in which Ohio practice aligns with best practice – such as Ohio Common Pleas judges’ near universal treatment of ARA risk scores as just one among many pieces of information to be considered when reaching a decision. It also identifies areas where Ohio practice falls short of best practice, such as in the training of judges, and on this basis generates the extensive list of recommendations set out in Section 5.

The recommendations provide a broad approach to improving the use of risk assessment tools in Ohio’s judicial system. They focus on several key areas: clear communication of results, comprehensive training for both assessors and judges, alignment with due process values, and robust oversight and accountability mechanisms. First, improving how risk assessment scores are presented to judges is essential. Courts should ensure that ARA scores are communicated clearly and consistently, emphasizing that these scores are just one factor among many in decision-making.

Additional research should also explore the best methods for presenting these results, whether through percentages, risk categories, or other visual aids.

Second, training is critical. Both assessors and judges require adequate, ongoing education to properly use and understand ARA tools. For assessors, this includes cultural sensitivity and unbiased interviewing techniques, while judges need specific training on interpreting risk scores and understanding the tools' limitations. Addressing disparities in training access, especially for rural judges or those new to the bench, is crucial to ensure consistent application across the state. In terms of due process, courts must provide judges with a better understanding of how risk assessments can affect legal outcomes and due process values. Consistent guidelines on how and when ARA scores are shared with defendants, along with sufficient education on their use, are necessary to ensure fairness and transparency.

Third, strong oversight is needed to ensure the accuracy and accountability of ARA tools. Independent evaluations, double-coding, and risk-based testing should be prioritized where feasible. Courts should also remain vigilant about external incentives that may skew risk scores, and they should regularly gather feedback from judges to refine ARA implementation. By adopting these recommendations, Ohio courts can enhance the fairness, transparency, and effectiveness of risk assessment tools, ensuring that they serve as a valuable aid in the judicial decision-making process rather than a determinant of outcomes.

Overall, and perhaps most importantly, this report shows that algorithmic risk assessments are deeply integrated into the Ohio Courts of Common Pleas. Judges largely respect them and use them. They are part of the fabric of criminal justice in Ohio. The scholarly and policy debate over whether courts should use algorithmic risk assessments should and will continue. But the fact that judges do use these tools shows the importance of understanding how they do so, and how they might do so more fairly, more transparently, and more in line with their own and Ohio's criminal justice values. This report seeks to contribute to that end.

References

- Bavitz, C., Bookman, S., Eubank, J., Hessekiel, K., and Krishnamurthy, V. (2018). Assessing the assessments: Lessons from early state experiences in the procurement and implementation of risk assessment tools. *Berkman Klein Center Research Publication*, (2018-8).
- Berk, R. A. (2021). Artificial intelligence, predictive policing, and risk assessment for law enforcement. *Annual Review of Criminology*, 4(1):209–237.
- Chanenson, S. L. and Hyatt, J. M. (2016). The use of risk assessment at sentencing: Implications for research and policy. *Hyatt, JM & Chanenson, SL (2016). The Use of Risk Assessment at Sentencing: Implications for Research and Policy. Bureau of Justice Assistance, Washington, DC.*
- Chohlas-Wood, A. (2020). Understanding risk assessment instruments in criminal justice. *Brookings Institute*. Retrieved from <https://www.brookings.edu/articles/understanding-risk-assessment-instruments-in-criminal-justice/>.
- DeMichele, M., Comfort, M., Barrick, K., and Baumgartner, P. (2021). The intuitive-override model: Nudging judges toward pretrial risk assessment instruments. *Fed. Probation*, 85:22.
- Doleac, J. L. and Stevenson, M. T. (2020). Algorithmic risk assessments in the hands of humans.
- Eckhouse, L., Lum, K., Conti-Cook, C., and Ciccolini, J. (2019). Layers of bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior*, 46(2):185–209.
- Garrett, B. L. and Monahan, J. (2020). Judging risk. *Calif. L. Rev.*, 108:439.
- Hamilton, M. (2020). Risk assessment tools in the criminal legal system—Theory and Practice: A Resource Guide. NACDL - National Association of Criminal Defense Lawyers. Retrieved from <https://www.nacdl.org/Document/RiskAssessmentReport>.
- Hill, S. A. I. (2021). Bail reform and the (false) racial promise of algorithmic risk assessment. *UCLA L. Rev.*, 68:910.
- Kansas v. Walls (2017). No. 116,027, slip op. at 6-7. Kan. Ct. App.
- Kuckes, N. (2006). Civil due process, criminal due process. *Yale Law & Policy Rev.*, 25:1.
- Latessa, E., Smith, P., Lemke, R., Makarios, M., and Lowenkamp, C. (2009). Creation and validation of the ohio risk assessment system: Final report. Retrieved from <https://www.uc.edu/content/dam/uc/ccjr/docs/reports/project3emORAS3em3em.pdf>.
- Latessa, E. J., Lovins, B., and Lux, J. (2018). The ohio risk assessment system. *Handbook of Recidivism Risk/Needs Assessment Tools*, pages 147–163.
- Mayson, S. G. (2018). Bias In, Bias Out. *Yale Law Journal*. A Survey of Pre-trial Practices in Ohio. <https://www.supremecourt.ohio.gov/docs/Boards/Sentencing/>

[resources/pretrialPractices.pdf](#).

Ohio Criminal Sentencing Commission (2020). A Survey of Pretrial Practices in Ohio. Retrieved from <https://www.supremecourt.ohio.gov/docs/Boards/Sentencing/resources/pretrialPractices.pdf>.

OSBF (2020). Balancing bail reform. *Ohio Bar*. Retrieved from <https://www.ohiobar.org/member-tools-benefits/practice-resources/practice-library-search/practice-library/20200L/balancing-bail-reform-criminal-rule-46/>.

Partnership on AI (2019). Report on algorithmic risk assessment tools in the U.S. criminal justice system. *PAI*. Retrieved from <https://partnershiponai.org/paper/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/>.

State v. Loomis (2016). 881 n.w.2d 749, Wisconsin Sup. Ct.

Stevenson, M. (2018). Assessing risk assessment in action. *Minn. L. Rev.*, 103:303.

Task Force (2019). Report and recommendations of the supreme court of Ohio Task Force to examine the Ohio bail system. *The Supreme Court of Ohio*. Retrieved from <https://www.supremecourt.ohio.gov/docs/Publications/bailSys/report.pdf>.

United States Probation (n.d.). Presentence Investigation. *Southern District of Ohio*. Retrieved from <https://www.ohsp.uscourts.gov/presentence-investigation>.