



**SUMMARY REPORT:
THE PRETRIAL PREDICTIVE VALIDITY OF THE
PUBLIC SAFETY CHECKLIST (PSC) &
VIRGINIA PRETRIAL RISK ASSESSMENT INSTRUMENT (VPRAI)
IN NINE OREGON COUNTIES**

Submitted to:

Oregon Criminal Justice Commission,
National Criminal Justice Reform Project,
& Arnold Ventures

Principal Investigator:

Christopher M. Campbell, Ph.D.
Department of Criminology and Criminal Justice
Portland State University
ccampbell@pdx.edu

Co-Investigators

Kris Henning, Ph.D.
Arynn Infante, Ph.D.
Ann Leymon, Ph.D.
Mark (Harmon) Leymon, Ph.D.
Mauri Matsuda, Ph.D.
Brian Renauer, Ph.D.
Curt Sobolewski, Ph.D.
Katie Wuschke, Ph.D.

Funded by:

Oregon Criminal Justice Commission,
National Criminal Justice Reform Project,
& Arnold Ventures

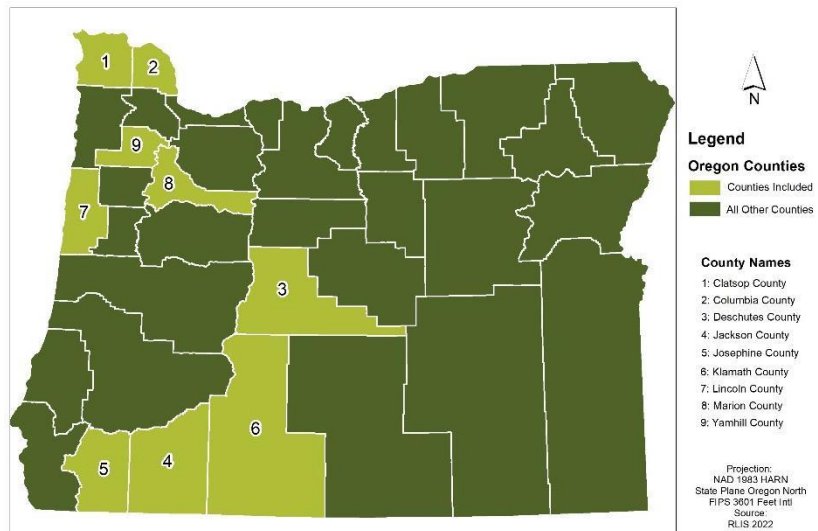
Date Completed: 6/16/2022

Background

Within the context of Oregon's Justice Reinvestment Initiative (JRI), the state has made concerted efforts to reduce the frequency and length of offender prison sentences. Given the nationwide concerns that the overuse of pretrial detention may result in more severe sentencing outcomes for criminal defendants, several Oregon counties have expressed a desire to reform their use of pretrial detention. In an effort to examine the decision-making process and influence of pretrial detention on prison sentencing, the Criminal Justice Commission (CJC) contracted Portland State University researchers to conduct an array of qualitative and quantitative analyses of Oregon's pretrial processes, programming and outcomes. The findings from the analyses uncovered the importance of pretrial risk tools in decisions to release or detain defendants (see Campbell et al., 2021, 2022). Specifically, the investigations on county- by-county pretrial decisions and protocol, emphasized the fact that jail capacity and operations routinely factor into decisions to release people on pretrial supervision. For instance, when jail capacity is a problem (i.e., emergency releases) the sheriff's office typically relies on the locally used risk tool to help create the grid to release people of the lowest risk. In the handful of jurisdictions that are relying on a risk tool to make both types of release decisions, the counties have expressed that the tool they are using has not been and needs to be validated, in spite of their reliance on it.

While some counties use a pretrial-specific tool such as the community supervision-specific Public Safety Checklist (PSC, a static, post-conviction risk tool), others use the Virginia Pretrial Risk Assessment Instrument (VPRAI) for release decisions and emergency release matrices specific to pretrial holds. The validation of the PSC and the VPRAI (i.e., ensuring the tools being used to make release decisions are accurate in predicting risk) is

critical for release decisions at the jail, pre-arraignment, and arraignment decision-points. Validating a risk scale is key to ensuring decision-makers are supplementing their professional judgement and maximizing community safety. Moreover, without assessments of accuracy, there is no way of determining if the tool or discretionary processes are associated with degrees of disparity that may exist across race/ethnic groups.



This project provides an evaluation of the predictive accuracy of the pretrial risk assessment tool used in nine counties Clatsop, Josephine, Marion, Jackson, Klamath, Columbia, Deschutes, Lincoln, Yamhill, and answers the following research questions:

1. How accurate are the pretrial risk tools used in when it comes to predicting pretrial outcomes (i.e., failure to appear, re-arrest)?
2. How can the risk tools and/or assessment protocol be changed to improve predictive accuracy?

None of the nine counties are linked to study findings by name. Rather, each county is referred to using a pseudonym (e.g., County A) to preserve the confidentiality of county-specific findings while allowing for broader findings and conclusions to be addressed in this report. County specific findings and potential solutions for addressing predictive accuracy problems are being shared directly with county officials.

Methodology

The data used in this validation study was compiled by merging information from the Oregon Judicial Department (OJD) data management system (Odyssey) with arrest information from the Law Enforcement Data System (LEDS). Steps used to complete this merge is detailed in the Appendix and is the same regardless of the county being studied. Measures used in this analysis were confined to the risk score in question, and other measures known to release authorities at the time of release. For information on measures used, definitions, and codes, please see Table B in the Appendix of this report.

Measures

Outcomes. The primary measures emphasized in this analysis, are three outcomes used to test the predictive validity of the risk tool. Each of the measures reflect a central concern of the courts when determining pretrial release: Failure to appear, a rearrest for any crime, and a rearrest for a person-related offense. Failure to appear (FTA) captures any mention of a failure by the defendant to appear for a court date at any time during the pretrial period. Rearrest for any offense¹ and rearrest for any person (i.e., violent) crime capture the breadth and severity of defendant behavior during the pretrial period. Each of these are coded as dichotomous (binary) events. Index cases were identified via a casefile date between 1/1/2018 and 12/31/2018. Dates related to release, hearing, and failure to appear were identified on these cases up to 9/1/2021 with the overwhelming majority of cases being disposed well before then. All rearrest data was pulled up to 1/20/2022 so long as the rearrest fell within the index pretrial period (casefile date and disposition).

Public Safety Checklist (PSC). There are three specific PSC scales designed to predict three separate recidivism outcomes for those convicted and serving a period on community supervision – (1) A new felony conviction, (2) a new property-related arrest, and (3) a new person-related arrest. Each scale consists of 31 items, 14 of which are interactions between two measures. As discussed on the Criminal Justice Commission’s website,² the PSC is an automated tool that checks the following static risk factors for each person entered and computes a level of risk for future crime:

1. age;
2. gender;
3. age of first arrest;
4. severity of current crime;
5. number of prior arrests for a person, property or other (statutory) type of crime;
6. was this person arrested for a person, property or other type of crime in the last five years;
7. any prior theft conviction;
8. any prior revocations on parole or probation;
9. any prior incarcerations;
10. multiple custody cycles;
11. previous sentence type (probation or incarceration).

Virginia Pretrial Risk Assessment Instrument (VPRAI). The VPRAI was originally designed for the Virginia Department of Criminal Justice Services to help determine a defendant’s likelihood of

¹ When possible and relevant, the “any arrest” measure removed any possession of a controlled substance arrest. Drug possession was classified as a violation (akin to a traffic citation) with the 2020 passage of Measure 110.

² Oregon Criminal Justice Commission website on the Public Safety Checklist risk tool:
<https://www.oregon.gov/cjc/risktool/Pages/default.aspx>

failing to appear or being rearrested during the pretrial period (VanNostrand, 2003). There are eight items used in the VPRAI, each given a point value that represents the item’s importance in predicting the outcome which can sum up to 9 points (see Appendix Table C).³ The instrument was designed to be part of an interview within the pretrial investigation report.

In Oregon, this is the common mode of collection, but only three of the items overtly require collection via speaking to the defendant – Length at current residence, employment status, and history of substance abuse. The rest can largely be collected by examining the individual’s criminal history provided by the Odyssey or LEDS systems, but these systems only provide the person’s history within the state. Having any points for a measure indicates that the measure is true for that person. Ultimately, when the point values are added together, the total is split to up to five groupings delineating defendant risk to fail pretrial supervision: 1 (Low risk), 2, 3 (Average), 4, and 5 (High risk). Counties using the VPRAI identified their own cut-points that did not necessarily coincide with these same levels. We used the county-defined cut-points as it reflects how the tool is used in practice. *It is important to highlight that while the VPRAI is the only pretrial-specific instrument being implemented among these counties, it has not yet been validated in these counties.* It is also worth noting that there is no set standard for setting risk-assessment cut-points.

Analysis

The analysis used to test the predictive validity of the PSC and the VPRAI included multiple components. First, a sensitivity and specificity analysis was conducted to assess the baseline performance of the tool. This involves testing the receiver operating characteristic area under the curve (AUC), which assesses the accuracy of the scoring system in predicting a true occurrence of the outcome. In other words, the AUC provides an indication as to how well the risk score can differentiate between someone who fails supervision and someone who is successful (i.e., returns to court and/or is not rearrested). The AUC statistic ranges from 0 to 1, with 1 indicating perfect prediction (100%) and .5 representing chance (i.e., equal to a coin flip or 50% accuracy). A tool’s AUC can be interpreted in the following ways according to prior literature (Rice & Harris, 2005):

AUC value thresholds	Interpretation
AUC ≥ .714	Strong
.714 > AUC ≥ .639	Moderately strong
.639 > AUC ≥ .556	Relatively weak
AUC < .556	Poor, unreliable

While all actuarial risk tools aim to have the highest accuracy possible, it is generally considered that most tools equal to or greater than .639 are often more accurate than subjective, professional judgement alone.

Second, it is also important to know if the risk tool can predict the outcome of interest (FTA or rearrest) while controlling for other factors that may be important or influential in how a defendant succeeds. To test the tool’s accuracy in this context, a series of logistic regressions⁴ were used to analyze

³ The original construction of the VPRAI (VanNostrand, 2003) included nine items totaling 10 points. A revalidation of the VPRAI (VanNostrand & Rose, 2009) removed the “outstanding warrants” measure, resulting in the current scale.

⁴ Logistic regression models are reported for each of the outcomes assessed. Each model reported include the bootstrapped standard errors replicated between 50 and 100 times, which ensures greater accuracy in determining the statistical significance of a measure within the model, particularly if there is a smaller sample size. Additionally, for situations in which there were low base rates of events (less than 20%) within small samples (less than 200), a

the ability of the risk tool to predict the outcome while accounting for other factors that may not be accounted for in the risk score itself (see the Appendix Table B). If the tool remains a strong, significant predictor of the outcome while other measures are controlled for, then the tool's predictive validity is relatively robust given these other factors are present.

Third is testing the tool's relative accuracy related to demographic groups. As detailed further in recent reports to the CJC (Campbell et al., 2022), recent attention on actuarial risk tools in the justice system has focused on the potential of inherent bias that exists due to the measures used in the risk assessment. Although the overarching goal at every stage of the justice system ought to be (and often is) to maximize accuracy and fairness when using risk assessments, this is not always the result. In order to identify and reduce bias, risk tools must be tested and retested for issues in accuracy and misclassification. Ideally, if a risk tool is validated and accurate for all race/ethnic populations, the tool is expected to be between roughly 64% to 70% accurate (AUC = .64 to .70, respectively), minimizing the potential for misclassification. For example, misclassification arguably occurs when an individual is identified as "high risk" but is actually "low risk" to recidivate. If bias is detected (e.g., clear misclassification for one group compared to another) then statistical changes can be made to correct them before the tool is used to help make decisions. Of particular concern is the presence of racial/ethnic bias in early subjective decisions (such as pretrial release), even if slight or "non-significant" in research findings, may have "cumulative disadvantages" across the justice system decisions and over time for defendants (Kramer & Wang, 2019; Stolzenberg et al., 2013; Wooldredge et al., 2015). For the purposes of this analysis, we examine race/ethnicity so long as there is a large enough sample to do so in the county. If the risk tool's AUC is statistically similar across racial/ethnic groups, then it is understood to have equal predictive accuracy regardless of the grouping to which it is applied.⁵ Unfortunately for many of the VPRAI samples, we were not supplied demographic information by the county, which rendered this third analysis impossible.

Sample Characteristics

Table 1 provides a breakdown of the descriptive statistics of all nine counties. While each of the counties began with several hundred or thousand cases, many cases were lost due to limiting the dataset to cases that could be linked to other databases (i.e., between Odyssey and LEDS), as well as due to needing to focus only on cases that were actually released, and therefore at risk to fail in the community. County sample size ranged from 285 to 2,314 for PSC analyses, and for the VPRAI samples ranged from 173 to 2,658 cases.⁶ Overall, the sample was overwhelmingly male (73.9%) and White (76.7%). The most serious charges included similarly proportioned property (24.6%), driving while intoxicated/under the influence (21.8%), and person crimes (19.2%).

The risk scores ranged the minimum and maximum values of the respective tools. Regardless of the specific scale, the PSC ranges from 0 to 1, essentially as a predicted probability of a given recidivism outcome. It is important to note that the PSC is designed and left as a probability so that counties, the

penalized regression was used to allow for a predictive model (Firth method; Firth, 1993; Heinze & Schemper, 2002) that would otherwise be biased or struggle with complete separation.

⁵ The PSC is designed to be a uniform scale for males and females. With the passage of Measure 110, there is an ongoing effort to re-design the PSC and potentially split the scale into a male- and female-specific score. Taken together, the aim to test and scrutinize male and female differences for a score that already accounts for sex was viewed as unhelpful and thus was not included in this report or the county-based reports.

⁶ Four counties were initially included in the VPRAI analyses. The fourth county, we will call County H, was removed due to the fact that the county only had 60 cases that met the eligibility criteria noted. It was also removed from the VPRAI analysis because the implementation of the VPRAI has been modified in this jurisdiction and is not comparable to other VPRAI implementation efforts examined here.

Department of Corrections, and the Criminal Justice Commission can determine the best cut-points for low, medium, and high risk based on capacity and other determinations about probation and post-prison supervision populations (for which the scale was designed). Scale cut-points are provided here to demonstrate how the scales are typically used for community corrections populations. Cut-point percentages determined by the Criminal Justice Commission are provided for the PSC-Felony scale as a point of reference. The largest proportion of the sample falls into the low-risk category (54.1%).

For the VPRAI, it is worth reiterating that these three counties were the only counties with a large enough sample size to analyze. Additionally, these counties reported that their implementation protocol is fidelitous to how the VPRAI was designed to be collected. That being said, there were notable differences in the baseline examination of the counties. Specifically, County I possesses a notable positive skew, with many more low scores. Meanwhile, County G was the only county with total point totals that reached the maximum 9-points possible on the VPRAI. Finally, the sample possesses sizable base rates for FTA (38.0%) and any arrest (36.9%) outcomes, while person arrests during the pretrial period were substantially lower (12.9%).

Table 1. Descriptives

	n	% or Mean (SD)	Min	Max
Entire sample	9,573		285	2,314
Demographics				
Male	7,076	73.9%		
Race/Ethnicity				
White	7,346	76.7%		
Black	295	3.1%		
Hispanic	1,534	16.0%		
Native American/Pacific Islander	145	1.5%		
Average age	24,437	34.2 (11.1)	16	88
Average age at first arrest	24,423	27.0 (10.7)	14	87
Most Serious Index offense				
Property	2,357	24.6%		
Motor Vehicle - DWI/DUI	2,081	21.8%		
Person	1,833	19.2%		
Domestic Violence Related	874	9.1%		
Drugs	987	10.3%		
Public Order	530	5.5%		
Felony - Weapon	337	3.5%		
PSC Information				
PSC-Felony	9,573	.30 (.22)	.01	.99
Low	5,179	54.1%		
Medium	1,999	20.9%		
High	2,395	25.0%		
PSC-Property	9,567	.24 (.19)	.01	.96
PSC-Person	9,567	.37 (.15)	.01	.96
VPRAI Score				
County B	173	3.14 (2.02)	0	8
County G	1,469	4.17 (2.22)	0	9
County I	2,658	.79 (1.71)	0	8
Outcomes				
FTA	3,641	38.0%		
Any Rearrest	3,534	36.9%		
Person Rearrest	1,231	12.9%		

Summary of Results

Overall Performance

With each of the counties having their own, in-depth analysis and reporting, this summary provides a brief breakdown of the overall performance of both the PSC and the VPRAI. Table 2 shows the baseline AUC values for each risk scale in its ability to predict the given pretrial outcome. As shown, the PSC-Felony and PSC-Property scales perform rather well in predicting FTA outcomes and any arrest outcomes for seven of the nine counties assessed. The PSC-Person scale yields a weak performance at best in predicting any outcome, including a rearrest for person crimes. In contrast, non-validated VPRAI application yielded a weak to poor performance for each outcome.

Most of the AUC values shown here suggest that the scales need further study and calibration. For example, in County B the implementation of any of these scales should be approached with extreme caution. For this county, only one scale, PSC-Person, performs well enough in predicting one outcome (person crime rearrest) to be relied upon in helping to make release decisions. On the opposite end of the performance spectrum, the PSC-Property performs rather well in County H and I for the outcomes of FTA and any rearrest. This bodes well for these counties that rely on the PSC to inform their release decisions.

Surprisingly, the VPRAI yields a weak performance at best in the three counties that implement it. This is likely due to the fact that the VPRAI is being used as an off-the-shelf tool for all outcomes in these counties, with the assumption that the predictive validity of the tool is the same in Oregon counties as it is in Virginia without changing the item point values. For example, the original point value for the item indicating that the “defendant has two or more FTAs” is 2-points, compared to 1-point for all other items. This point value was determined by identifying its importance in predicting pretrial outcomes (FTA and committing a new crime) in the Virginia sample (VanNostrand & Rose, 2009). It is possible that this same item may have more or less weight in the ability to predict the outcomes important to Oregon counties. The only way to know, is by re-validating the VPRAI and calibrating it on Oregon populations. Individual reports to the counties will begin to examine this possibility.

Table 2. Baseline AUC statistics predicting pretrial outcomes

Risk Scale	PSC - Felony			PSC - Property			PSC - Person			Non-validated VPRAI		
	FTA	Any	Person	FTA	Any	Person	FTA	Any	Person	FTA	Any	Person
County A	.686	.680	.581	.690	.691	.567	.545	.570	.569	-	-	-
County B	.528	.545	.455	.554	.580	.527	.483	.531	.628	.579	.560	.548
County C	.663	.660	.581	.686	.691	.631	.523	.537	.616	-	-	-
County D	.649	.642	.547	.701	.697	.603	.516	.513	.618	-	-	-
County E	.650	.648	.563	.695	.701	.628	.550	.578	.609	-	-	-
County F	.578	.582	.451	.664	.662	.572	.567	.568	.600	-	-	-
County G	.648	.650	.557	.697	.692	.600	.558	.570	.634	.627	.636	.564
County H	.705	.719	.579	.734	.738	.632	.595	.594	.617	-	-	-
County I	.671	.675	.553	.717	.705	.593	.564	.539	.597	.624	.614	.555

As Table 2 shows, the AUC performance of each scale appears to run the gamut in terms of accuracy, and varies widely across the different counties. While this is a critical metric when it comes to assessing a risk assessment’s performance, it is not the only metric. One of the key tasks related to pretrial risk tools in particular, is identifying the low-risk individuals. Identifying low-risk defendants highlights those who are least likely to commit a new offense during the pretrial period and are most likely to return for their next court date. Thus, a scale, and particularly the cut-point scheme that can identify a low risk

population that is (1) significantly different from higher-risk populations, and (2) yield a rather low percentage of those who fail during the pretrial period. In other words, a good risk assessment is one that clearly distinguishes low-risk individuals from those who are higher-risk, and accurately identifies people who are actually low-risk.

Table 3 shows the percent of defendants who failed during the pretrial period who were identified as “low-risk” under the given scale, while still shaded in accordance with the AUC performance from Table 2. Some overlap between the tables is clear. For example, County H yielded a strong AUC performance (.719) while also significantly differentiating the low- from high-risk defendants ($p < .001$), and having a relatively lower percentage of defendants who fail by way of any rearrest actually classified as “low-risk” (31.4%). Many if not most of the poor AUC performing scales could not differentiate low- from high-risk defendants ($p > .05$), and had a much higher percentage of defendants failing during the pretrial period while also classified as low-risk (near 50% or higher). This is particularly problematic with the VPRAI in County B and I which yielded the higher percentage of FTAs and rearrests that fell into the low-risk classification. Perhaps even more problematic are the scales with moderately strong AUC values that also have a high percentage of defendants who fail/recidivate and are classified as “low-risk.” These scales likely need an adjustment of the associated cut-points to help better distinguish those defendants who are actually lower risk to FTA or be arrested for a new crime.

Table 3. Percent of defendants who failed during the pretrial period and were low risk

Scale	PSC - Felony			PSC - Property			PSC - Person			Non-validated VPRAI		
	FTA	Any	Person	FTA	Any	Person	FTA	Any	Person	FTA	Any	Person
County A	38.2***	40.0***	59.3	50.6***	52.2***	63.0	66.3	73.3	55.6	-	-	-
County B	52.1	48.7	60.5	61.7	59.1	65.1	78.7*	72.2	62.8	33.3	41.0	35.0
County C	47.9***	46.7***	63.2	59.2***	57.6***	75.5	70.9	68.8	61.4**	-	-	-
County D	35.3***	30.9***	44.2	47.1***	42.8***	58.5	75.1***	75.9***	59.5***	-	-	-
County E	32.6***	29.5***	39.5	44.1***	39.7***	46.5*	84.7*	84.5*	79.0	-	-	-
County F	53.9	40.1***	48.8	58.3**	52.9***	60.5	78.9**	79.0*	69.8	-	-	-
County G	42.5***	37.5***	48.4	49.3***	48.8***	58.1	76.5*	77.0**	63.4	7.1***	7.1***	7.6**
County H	37.3***	31.4***	42.7***	45.6***	41.2***	54.3***	75.2***	73.9***	62.1	-	-	-
County I	37.3***	33.2***	50.0	47.9***	45.6***	55.4	74.2*	77.0***	62.5	26.2***	22.8***	22.2**

* $p < .05$; ** $p < .01$; *** $p < .001$

Race/Ethnicity Group Performance

Due to the low counts of different race/ethnic groups, all assessments of PSC performance by race were completed via a White versus Non-White differentiation.⁷ Table 4 presents the AUC statistics for each PSC scale by each pretrial outcome within the racial grouping. The AUC values are color coded to coincide with the AUC strength identified by prior research (Rice & Harris, 2005). A problematic discrepancy between the racial groups can be understood as any meaningful difference in the deviation of accuracy for one group over another. The definition of “meaningful” is difficult to operationalize. Using p values or tests of significance, for instance, is often too dependent on sample size and may be misleading. Arguably, a discrepancy is problematic if the difference in AUC value (i.e., accuracy) between the groups is more than 5% (i.e., the difference in AUC values between .650 and .700). This is especially the case if the difference spans a strength threshold (i.e., color code).

⁷Demographic information was not acquired for the VPRAI prior to the drafting of this report. We are in the process of securing that data and are attempting to incorporate it into follow-up county-specific analyses.

As shown in Table 4, regardless of the scale, there is some kind of problematic discrepancy observed in every county. Some predict a given outcome equally well. For example, in County H, the PSC-Felony and PSC-Property both predict a rearrest for any crime equally well for both racial groupings. In several other counties, there are instances in which a scale may predict the outcome equally poorly. For County C, for instance, the PSC-Property scale predicts rearrests for person crimes poorly. Overall, depending on the scale being used, and the county it is used in, there is indeed a concern that a scale may be producing biased results when it comes to performance among defendants of differing racial groups. This lends further support for the notion that all scales being used to aid in making pretrial release decisions need to be tested and validated to the specific jurisdiction and population served.

Table 4. AUCs predicting pretrial outcomes by White versus Non-White groups

County		PSC - Property			PSC - Felony			PSC - Violent		
		FTA	Any	Person	FTA	Any	Person	FTA	Any	Person
A	Non-White	.713	.831 ^a	.736 ^a	.707	.841 ^a	.386	.716 ^a	.707 ^a	.686 ^a
	White	.675	.671	.547	.683	.667	.465	.564	.552	.554
B	Non-White	.605 ^a	.723 ^a	.565 ^a	.530	.713 ^a	.399	.364	.457	.511
	White	.540	.563	.526	.527	.538	.532	.470	.539	.642 ^a
C	Non-White	.591	.675	.534	.603	.662	.483	.496	.645 ^a	.663 ^a
	White	.673 ^a	.694	.538	.676 ^a	.693	.477	.598 ^a	.630	.606
D	Non-White	.626	.695	.559 ^a	.633	.684	.451	.515	.615	.606
	White	.644	.697	.503	.651	.704	.492	.554	.601	.622
E	Non-White	.699	.676	.698 ^a	.719 ^a	.658	.333	.542	.624	.706 ^a
	White	.640	.703	.558	.640	.701	.468	.566 ^a	.629	.594
F	Non-White	.553	.653	.722 ^a	.567	.684	.263	.423	.552	.678
	White	.587 ^a	.668	.506	.579	.661	.501	.457	.581	.564
G	Non-White	.756	.728	.709 ^a	.758 ^a	.722 ^a	.319	.605 ^a	.645 ^a	.727 ^a
	White	.630	.685	.526	.626	.691	.479	.547	.588	.596
H	Non-White	.704	.746	.613	.683	.740	.389	.575	.637	.608
	White	.728 ^a	.732	.581	.720 ^a	.731	.416	.585	.632	.624
I	Non-White	.645	.676	.573 ^a	.642	.700	.456	.535	.593	.582
	White	.686	.709	.530	.682	.719 ^a	.358	.558 ^a	.596	.605

a = Problematic discrepancy.

It is worth noting that when broken out by racial groupings, the PSC-Felony scale produces a biased performance when predicting rearrests for person crimes. Similarly, the PSC-Violent scale produces rather poor predictions for FTAs. With both scales producing AUCs less than .5, this indicates that they are better at predicting the null (a non-event) than they are at predicting the recidivism event. This is a testament as to how specialized these scales are in their predictive performance and how they were meant to be used.

Limitations

There are several limitations to consider when interpreting the results of these analyses. First, as with all jurisdictions across the nation and within Oregon, the FTA measure is rather volatile. As discussed in the prior two reports on Oregon pretrial processes (Campbell et al., 2021, 2022), counties can widely vary by how they approach the application of FTA charges and warrants, and even how they are captured in the hearing cancellation data. These factors have the potential to impact the predictive validity of the tool in some way. We attempt to account for this by using a rather broad estimate of the number of FTAs a

defendant may experience, as it captures any mention of whether the defendant failed to appear for a hearing at any time during the duration of the case (case file date to disposition).

A second limitation is the rearrest measure. As with any measure of arrest, there is the potential that it captures more systemic practices than the defendant behavior. Noted recently in publications focused on predictive bias in risk tools, criminal history and arrest measures in particular may be more indicative of over-policing practices and those that happen to target racial/ethnic minorities. Although we attempted to assess the degree to which the scales were predictive of the outcome by racial group, the potential is still germane. Analyses of predictive problems may suggest that the application and local police practices may need further analyses than what is provided across our reports.

Third, the counties varied widely by their sample size, which may impact the confidence of our assessment of a given scale. Sample size became a concern for three of the counties and was particularly an issue for the VPRAI. Three counties could not be included in the VPRAI analysis because their samples were less than 100 cases. Much of the small sample issues were due to the fact that there were many cases that could not be identified in both Odyssey and LEDS databases. Further efforts need to be engaged in to improve how the data systems can be linked using case and state identification numbers.

Fourth, one major issue identified by the prior two analyses (Campbell et al., 2021, 2022) is that of jail capacity. As jail capacity becomes a problem in a given county, the process by which defendants are released or detained is dramatically changed. When a jail reaches capacity, some counties opt to use a release matrix for only those adults in custody who are already convicted and serving a sentence in local control. Other counties opt to identify those pretrial defendants who are low risk and release them to some version of pretrial monitoring if it is available, and others are released on recognizance regardless of the bail/security that was set. When defendants are released due to capacity issues, it may supersede a judge's order and security set. Essentially, what was a judicial determination has the potential to become an executive decision of the sheriff's office. Ideally, we would be able to identify this population to track during the pretrial period and distinguish them from pretrial defendants who are released on recognizance or released on conditions of monitoring. In spite of this being an issue in some counties, we have little reason to believe that force/emergency releases due to jail capacity make up a large proportion of the cases analyzed, and thus are unlikely to impact the results overall; although it could be important for some.

Finally, the cases analyzed during this timeframe varied in their follow-up period. This is due to the fact that it is rather difficult to truncate a pretrial period as cases can linger for months depending on the type of case. One aspect that lends strength to our study is the codified expectation that all circuit courts should aim to resolve/dispose cases within 60 days or require a continuance is filed. Many if not most cases are disposed within or shortly after this window. Unfortunately, many of the cases analyzed spanned the duration of the pandemic and the 2020 shutdown. The stay-at-home order issued in March of 2020 marked the beginning of a reverberating backlog in the courts which was often associated with failure to appear. The OJD in most circuits was quick to adapt, however, in allowing call-in and video-conference hearings for defendants who were not detained pretrial. This allowed for most jurisdictions to continue with their normal, albeit slower, functioning.

Conclusions

As indicated by Table 2, the overall performance of the pretrial risk scales assessed here vary rather widely. There are indications that the tested scales perform rather well in some of the counties. In those cases, officials can take solace in the fact that the tool used provides important information to aid in their decision-making for pretrial release and even setting security. However, the performance of the tool highly depends on which tool is being relied on (e.g., PSC-Felony versus the PSC-Property or PSC -Violent).

Perhaps the more important issue is the fact that more often than not the scales analyzed performed rather poorly in predicting any outcome. Importantly, even with the poor prediction performance that are identified currently, these scales still largely provide better predictions than chance (a coin flip) and can help supplement a release authority's decisions.

It is critical to note, that the results presented here do not mean that the PSC and VPRAI are unreliable or inaccurate tools. On the contrary, they perform rather well in the situations and jurisdiction for which they were designed. Instead, their needs to be an effort to re-calibrate the different scales to make them more applicable to the population served and outcomes predicted. That said, these scales should be used with great caution and should be re-weighted. For the VPRAI counties, we are offering re-weighting suggestions for the counties to implement moving forward. For example, a basic effort to re-weight the same items used in the VPRAI for County B can improve the AUC from .579 for FTA outcomes to .683. This re-weighting and assessment can easily be employed on every county using the VPRAI. For those solely relying on the PSC, the county-specific reports along with the prior reports on pretrial processes will provide guidance with their specific use moving forward.

Finally, the context of these findings must be couched in past reports produced by the Oregon Public Safety Task Force,⁸ the Oregon Advisory Committee to the U.S. Commission on Civil Rights,⁹ Oregon Senate Bill 48, and the recent order put in place on June 7, 2022 by the Oregon Chief Justice. These reports and orders have laid the foundation to improve pretrial practices and risk tool use. Most notably, is the Chief Justice Order that requires all felonies and misdemeanors will be identified as falling in accordance with a designated "Release Guidelines Categorization List." In relation to risk assessments, specifically, the Chief Justice's Order on pretrial release orders (PRO, No. 22-010) states:

Each judicial district should identify and consider using a risk assessment tool to assist with release determinations, as provided in the judicial district's PRO. The PRO may direct the entity supervising the local correctional facility as to the appropriate use of the risk assessment score. A judicial district may use risk assessment tools to inform decisions about which type of release is appropriate for a defendant or in setting the conditions of release. If a judicial district uses a risk assessment tool, the Presiding Judge shall review the selected tool, to ensure that it is reliable and unbiased, and the selected tool shall be validated at least every five years or following significant changes to the population or laws and policies related to arrest/citation, detention, and sentencing. Chief Justice Order No. 22-010; Order establishing release guidelines governing presiding judge standing pretrial release order; pg. 6.

Our findings highlight the necessity and importance of this order. We cannot emphasize enough the importance of ensuring that jurisdictions use a tool to help make pretrial release decisions and that the tool is in fact empirically accurate in its ability to predict the key pretrial outcomes.

⁸ [Oregon Public Safety Task Force report per House Bill 2238 \(2017\) - December 4, 2020 | Oregon State Library](#)

⁹ [06-29-Pretrial-Release-Detention-and-Bail-Practices-Report.pdf \(uscrr.gov\)](#)

References

- Campbell, C., Henderson, K., & Renauer, B. (2021). *Report on examining pretrial detention in Oregon: A qualitative analysis of decision making*. Portland State University, Oregon Criminal Justice Commission. <https://doi.org/10.13140/RG.2.2.23157.29926>
- Campbell, C., Henderson, K., & Renauer, B. (2022). *Pretrial processes and sentencing outcomes in Oregon: Integrating quantitative and qualitative findings*. Criminal Justice Policy Research Institute, Portland State University. <https://doi.org/10.13140/RG.2.2.30885.58081>
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1), 27–38. <https://doi.org/10.1093/biomet/80.1.27>
- Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, 21(16), 2409–2419. <https://doi.org/10.1002/sim.1047>
- Kramer, K. L., & Wang, X. (2019). Assessing Cumulative Disadvantage against Minority Female Defendants in State Courts. *Justice Quarterly*, 36(7), 1284–1313. <https://doi.org/10.1080/07418825.2019.1685122>
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and Human Behavior*, 29(5), 615–620. <https://doi.org/10.1007/s10979-005-6832-7>
- Stolzenberg, L., D'Alessio, S. J., & Eitle, D. (2013). Race and Cumulative Discrimination in the Prosecution of Criminal Defendants. *Race and Justice*, 3(4), 275–299. <https://doi.org/10.1177/2153368713500317>
- VanNostrand, M. (2003). *Assessing risk among pretrial defendants in Virginia: The Virginia Pretrial Risk Assessment Instrument*. Virginia Department of Criminal Justice Services. <https://www.dcjs.virginia.gov/sites/dcjs.virginia.gov/files/publications/corrections/assessing-risk-among-pretrial-defendants-virginia-virginia-pretrial-risk-assessment-instrument.pdf>
- VanNostrand, M., & Rose, K. J. (2009). *Pretrial risk assessment in Virginia: The Virginia Pretrial Risk Assessment Instrument*. Luminosity, Inc. <https://www.dcjs.virginia.gov/sites/dcjs.virginia.gov/files/publications/corrections/virginia-pretrial-risk-assessment-report.pdf>
- Wooldredge, J., Frank, J., Goulette, N., & Travis III, L. (2015). Is the Impact of Cumulative Disadvantage on Sentencing Greater for Black Defendants? *Criminology & Public Policy*, 14(2), 187–223. <https://doi.org/10.1111/1745-9133.12124>

Appendix

Public Safety Checklist (PSC) Pretrial Risk Assessment Project – Methodology Summary

The following bullets are a summary of the steps taken to identify eligible cases in each county for the purpose of investigating the performance of the PSC in predicting pretrial outcomes.

1. CJC provided PSU with data set consisting of all criminal cases referred to Oregon’s Circuit Courts (i.e., Odyssey data system) in 2018. The file contained de-identified data on 84,692 cases.
2. CJC provided us with all of the LEDS data for each person (SID) in the Odyssey record set. We used these data to identify the arrest date for each Odyssey case. This included searching in LEDS for cases with the same SID -and- county of origin or citing agency. Arrest date in LEDS had to be exact match or within 6 days of either Case File Date, Offense Date, or First Pretrial Release Date from Odyssey.
3. CJC provided us with the full Odyssey records for each person (SID) in the 2018 case pool. This would include the index charges for the 2018 case, as well as all prior and future cases. This is truncated on the historical end to cases transferred into Odyssey when the data system started and ends in January 2022 when the data were pulled.
4. CJC provided us with the PSC risk factors and total scores for each case/defendant based on their Case File Date (i.e., what would the PSC score(s) be on the date the case was initially filed by prosecutors).
5. Cases were screened to identify a final sample that we were reasonably sure involved an actual arrest and jail booking -and- had a release to the community before the case was disposed.
6. Cases were eliminated from the pool for a variety of reasons that are documented below. Note that a case could be eliminated for multiple reasons.
 - 15,858 cases had a defendant that without a SID (state ID number). This could indicate that the person was not ever physically arrested and fingerprinted. Missing SIDs also prevented us from connecting the person to arrest dates in LEDS.
 - 27,024 cases we were unable to accurately identify an arrest date for the given case in LEDS. This could be due to data entry errors in Odyssey or LEDS, the person might have been cited and released never generating an arrest, they were missing an SID.
 - 31,233 cases were dropped because Odyssey did not list any release agreements for the defendant. Presumably these are cite and release cases rather than true physical arrests and jail booking.
 - 3,304 cases were dropped because the most serious offense listed was a violation rather than misdemeanor or felony. Violations are less likely to involve a physical custody and, as a result, agencies are less likely to screen them for pretrial release using the PSC.
 - 1,825 cases were dropped because the most serious offense listed was a Failure to Appear. The primary use of the PSC for pretrial decision-making is whether people charged with a new crime (i.e., other than FTA) are at risk for FTA or a new offense if released.
 - 6,396 cases were dropped because the sole charge listed was for drug possession. These offenses are much less likely to be of concern post legal changes to decriminalize drug possession.
 - 23,349 cases were dropped because the full Odyssey records for the given defendant (i.e., SID) were not found in the supplemental data CJC provided. This is largely due to the fact that some of the index cases from 2018 were missing an SID which I needed for linking purposes. Other index cases that were dismissed might also be unavailable.
 - 6,237 cases were dropped because there was an FTA issued at a hearing and this happened before the first release date. These cases probably result from citations in lieu of an arrest where the person fails to show up for their first hearing.

- 573 cases were dropped because their first release date preceded the arrest date we assigned to the case (see LEDS matching above). This probably results from inaccurate identification of the arrest date for the given case/charges.
- 1,962 cases were dropped due to other inconsistencies in the dates for various court processing decisions. For example, some arrest dates identified precede the offense happened date listed in Odyssey.
- 2,893 cases were dropped because the defendant was apparently held more than 90 days before their first pretrial release event. This raises concerns about the accuracy of the arrest date, release date, or both.

The factors listed above alone or in combination resulted in the loss of 58,679 cases (69.3%) from the initial pool. That left us with 26,013 cases (3.7%) where we are fairly confident that the defendant was arrested, booked into jail, and given a release agreement (i.e., pretrial release). From the reduced sample of cases we eliminated another 1,575 cases that were missing the scores on the PSC. Our final sample included 24,438 cases or 28.9% of the original pool. These cases were distributed as follows by county:

Table A. Observations by county

County of Record	# cases	County of Record	# cases
Baker	128	Lake	60
Benton	583	Lane	850
Clackamas	1,593	Lincoln	658
Clatsop	285	Linn	839
Columbia	432	Malheur	407
Coos	238	Marion	2,314
Crook	355	Morrow	29
Curry	179	Multnomah	3,234
Deschutes	1,334	Polk	739
Douglas	495	Sherman	23
Gilliam	18	Tillamook	297
Grant	87	Umatilla	547
Harney	90	Union	242
Hood River	168	Wallowa	23
Jackson	2,028	Wasco	176
Jefferson	217	Washington	3,247
Josephine	1,204	Wheeler	1
Klamath	611	Yamhill	707
Grand Total		24,438	

LEDS data and Odyssey hearing and warrant data were then used to code “recidivism” variables that reflect problem behavior during the defendant’s time on pretrial release (i.e., the days between the first release event and case disposition). This included any evidence of an FTA for a court hearing, a new arrest for a criminal offense, and a new arrest for a person crime as defined by CJC.

The following provides a breakdown of the measures used in this analysis. Any additional measures used by the researcher beyond this list will be indicated in the methodology section.

Table B. Measures used in the validation study

Variable	Description
Demographic Factors	
Male	PSC - Coded '1' if the defendant was male.
Race/ethnicity	Race/ethnicity of the defendant. – Dichotomized as White = 0 vs Nonwhite = 1 in some analyses due to sparse observations
Age	PSC - defendant's age at time of index release.
Index Offense Information	
DV-related	Dichotomous flag available in Odyssey indicating if the case involves domestic violence.
Offense degree	Identifies the type of criminal charge for the most serious offense (e.g., Misd A, Felony B).
Offense type	UCR crime type for the defendant's most serious charge (e.g., person, property, public order).
Time in community	Time at risk of recidivating in the community measured in weeks. This captures the time between the date the defendant was released from custody and either the disposition, FTA, or rearrest event whichever comes first.
Public Safety Checklist (PSC) Scores	
PSC-Felony	PSC - Combined risk score for new FELONY CONVICTION.
PSC-Property	PSC - Combined risk score for new PROPERTY offense
PSC-Violent	PSC - Combined risk score for new PERSON offense
PSC Risk Level	General cut-points applied to the PSC-Felony score: Low = PSC-Felony score less than .25 Medium = PSC-Felony score greater than or equal to .25, but less than < .42 High = PSC-Felony score greater than .42
Pretrial Outcomes During Pretrial Release Period (i.e., first release date to case disposition date)	
FTA	Coded 1 if the defendant had any type of FTA during the pretrial release period. Includes all three types of FTA listed below.
Any rearrest	Coded 1 if the defendant had any new arrest for a criminal offense in LEDS. Includes any type of criminal charge.
Person rearrest	Coded 1 if the defendant had any new arrest for a PERSON offense in LEDS

Table C. VPRAI information

VPRAI Items	Description	Point value
Charge type	The most serious charge is a felony.	1
Released pending trial	The defendant was on a form of pretrial release at the time of the alleged offense.	1
Criminal history	Adult criminal history includes at least one misdemeanor or felony conviction.	1
History of failure to appear	Defendant has two or more FTAs.	2
History of violent convictions	Defendant has two or more violent convictions.	1
Length at current residence	Defendant has lived at his or her current residence for less than one year or does not have a stable residence.	1
Employed, primary caregiver, student, retired or disabled	Defendant was not employed, a primary caregiver, student, retired, or disabled at the time of arrest.	1
History of substance dependence or abuse	The defendant has a history of substance dependence or abuse.	1

Table adapted from VanNostrand (2003).