

Riverside Pretrial Assistance to California Counties (PACC) Project

Validation of a Pretrial Risk Assessment Tool

Report Submitted by:

Brian Lovins
brian.lovins@uc.edu

Lori Lovins
lori.lovins@uc.edu

Correctional Consultants Inc.

Initial Report: November 30, 2015

Final Report: May 3, 2016

Table of Contents

Executive Summary	3
Project Introduction	5
Pretrial Risk Assessment Tools	5
Study Methods	6
Sample.....	6
Variables	6
Analysis.....	6
Sample.....	7
Validation Results for the Virginia Pretrial Risk Assessment Instrument (VPRAI)	8
Risk Score and Pretrial Failure	8
Risk Level and Pretrial Failure	9
Modified Risk Level Categories	9
Examining the VPRAI by Subpopulations	9
Beyond the VPRAI: A Revised Pretrial Risk Assessment Tool.....	11
Implementation of the Modified Pretrial Tool.....	17

Executive Summary

Correctional Consultants Inc. (CCI) was contracted by the Crime and Justice Institute to assist Riverside County Probation Department (RCPD) in the validation of a pretrial risk assessment tool. The RCPD Pretrial Services Unit was established to assist the court in making decisions regarding defendant releases and to monitor defendants in the community to compliance with pretrial conditions. The analysis presented in this report is based on data collected by RCPD between April 2014 and October 2015. The sample that was analyzed included 568 defendants whose pretrial supervision had been terminated. Seventy-two percent were males and 63 percent were persons of color. The full report provides a brief background on the use of pretrial assessments as well as a description of the data collected by RCPD using the Virginia Pretrial Risk Assessment Instrument (VPRAI) and predictive validity of the VPRAI for Riverside's population. Also included are recommendations for a locally validated pretrial risk assessment tool based on Riverside's own data as well as implementation guidelines.

Analysis confirmed that the VPRAI is able to separate pretrial failure rates by risk score; as risk score increases, so does the pretrial failure rate. However, the failure rate between risk categories was not significantly different between the VPRAI categories of low risk (9.5% failure) and below average risk (12.2%). The existing risk levels also resulted in similar failure rates for average risk (28.6%) and above average risk (30.6%) categories. The high risk failure rate was 38%. Adjusting the cutoffs and creating three risk levels provided a clearer separation between low risk (11.6% failure), moderate risk (29.2%), and high risk (38%).

To determine the validity of the adjusted risk levels by gender and race, for the analysis compared results for males and females, and for Whites and people of color. The adjusted VPRAI classified relatively well for these subgroups, though males and people of color had higher failure rates among moderate risk defendants.

To determine if the current VPRAI can be modified to create a more accurate assessment tool based on Riverside's data, the analysis examined 16 different demographic, residential stability, and criminal history items. Five variables were found to be predictive when combined with other variables: pending charges at time of arrest; two or more FTA in past two years; substance abuse problem; number of previous adult convictions; and more than one year at current residence.

These five items were combined to create a risk score between 0 and 5. Similar to the original VPRAI, as the risk score increased, so did the failure rate. Three risk levels were created which classified individuals into distinct failure categories: low risk (13% failure), moderate risk (27.2%), or high risk (42.5%). The analysis also examined risk classification by race and gender and confirmed that the modified tool differentiates low, moderate, and high risk equally well for males and females, and for Whites and people of color.

Based on this validation study and best practices in risk assessment implementation, recommendations include the following:

- The Department should continue to collect data and repeat the validation analysis with a larger sample of women and modified data collection for employment status;
- If adopting the modified risk tool, the RCPD should revise policies and procedures to reflect changes to the risk tool and train all staff and stakeholders on the new tool;
- The RCPD should consider examining inter-rater reliability to ensure accurate and consistent scoring across staff members;
- The RCPD should continue to use strategies for release supervision based on level of risk;
- The department should continue to use the established override procedures; and
- The department should continue to use the Community of Practice Group to get periodic feedback from staff on risk tool scoring and implementation.

Project Introduction

Correctional Consultants Inc. (CCI) was contracted by the Crime and Justice Institute (CJI) to assist Riverside County Probation Department (RCPD) in the validation of a pretrial risk assessment tool. The Pretrial Services Unit was established to assist the court in making decisions regarding defendant releases. Moreover, the Pretrial Services Unit monitors defendants in the community and ensures compliance with pretrial conditions.

The RCPD has been working with CJI since 2012 to examine their pretrial process, including selection and validation of a pretrial tool to provide objective data to decision makers regarding pretrial releases. The use of a validated tool will provide RCPD with the ability to manage resources, protect the community, and ensure that defendants who are released are supervised based on risk levels. CJI, RCPD, and Correctional Consultants worked together for several months to determine the appropriate data points to be collected. RCPD implemented the data collection in April 2014 and supplied CCI with those data for a preliminary analysis in March 2015 and a second analysis based on additional closed cases in October 2015. The following report provides a brief background on the use of pretrial assessments as well as a description of the data collected by RCPD using the Virginia Pretrial Risk Assessment Instrument (VPRAI) and predictiveness of the VPRAI for Riverside's population. Finally, recommendations for developing and implementing a revised risk assessment tool are provided.

Pretrial Risk Assessment Tools

The purpose of a pretrial risk assessment tool is to assist courts in predicting the likelihood that a defendant will fail if released to the community before disposition of his or her case (Summers and Willis, 2010). Failure is typically determined by either failure to appear (FTA) for a scheduled court date or by arrest for criminal behavior prior to case disposition. Generally these tools examine both flight risk and risk for further criminal behavior. One goal of a pretrial risk assessment is to standardize recommendations about pretrial release so that these decisions are less subjective and more consistent (Coopriider, 2009). A second goal is to maximize the success rates of pretrial releases. This requires that a maximum number of defendants are released without compromising FTA rates or community safety (Summers and Willis, 2010). Notwithstanding issues of crowding and decreasing budgets, the presumption that defendants are assumed innocent until proven guilty underscores the importance of maximizing successful releases (Lowenkamp, 2008).

In a study of more than 500,000 cases processed through the Federal Pretrial Services System, several risk factors were deemed predictive of risk of pretrial failure: the nature of the pending charges, criminal history, community supervision at the time of arrest, history of FTA, history of violence, employment stability, residential stability, and substance abuse (VanNostrand and Rose, 2009; Winterfield, Coggeshall, and Harrell, 2003). Levin (2007) found that jurisdictions

that used quantitative pretrial risk assessment tools had lower FTA and re-arrest rates as well as fewer problems with jail overcrowding. Thus, more jurisdictions appear to be considering the use of pretrial risk assessments.

Study Methods

Sample

Data for the development of the RCPD pretrial risk assessment were provided to Correctional Consultants Inc. (CCI) for analysis.¹ The data used for this study were collected by field staff between April, 2014 and September, 2015. To ensure that data were collected consistently, RCPD, CJI, and CCI developed a user guide that clearly identified the data points along with a detailed scoring guide. A standardized data collection process was used and data were captured electronically.

Variables

The three outcome types tracked by RCPD—failure to appear, new arrest, and technical violation—were combined into one outcome variable of unsuccessful termination from the program for this analysis. In this analysis and report, ‘failure’ describes unsuccessful termination for any reason. A range of potential predictors were included in the analysis to determine what was predictive of failure. Table 1 below lists each of these variables.

Date of Birth	Race
Gender	Education Level
Current Military Status	Housing Stability/Homelessness
Current Offense	Current Charge Count
Previous offenses	Pending charges
Prior supervision history	Age at 1 st arrest
Prior FTAs	Previous Violent Offenses
Employment	Alcohol/Drug use
Current risk level	Termination type
Outcome	

Analysis

The validation of the VPRAI was completed in multiple stages. First, bivariate analyses were conducted to determine the variables correlated with any failure, failure to appear, and arrest for a new crime. Next, the variables that were identified through the bivariate relationship were combined in a stepwise logistic regression model to determine if any of the variables could be

¹ All identifying information was removed from the datasets to protect participant identity.

eliminated. In addition to examining the additive benefit of each variable, logistic regression analyses were conducted to ensure that the variables selected were not highly correlated with gender or race.²

Upon identifying the predictors of failure, each item was converted to a 0/1 scoring system similar to the Burgess Scale (Burgess, 1928). While recent risk assessments use a more robust weighting system, it was decided that the simplicity of a simple additive scoring process provided more face validity to the assessment while still providing a valid assessment of risk (Nuttall, Barnard, Fowles, Frost, Hammond, Mayhew, Pease, Tarling, & Weatheritt, 1977).

For each version of the tool examined in this analysis, Receiver Operating Characteristics (ROC) analyses were conducted to determine the specificity of the model. The ROC analysis measured by the Area Under the Curve (AUC) provides a measure of accuracy by balancing false positives with false negatives to determine how well the tool predicts over chance. An AUC of .500 would suggest that a tool was not able to classify a person any better than chance. Schwalbe (2007) found that AUCs ranged from .532 to .780 with an average AUC of .640 across 18 risk assessments examined.

Sample

Table 2 summarizes demographic characteristics of the sample used for this study. Overall, there were 568 defendants who completed pretrial services as of October, 2015. Of the sample, 408 defendants were male and 160 were female. Nearly 64 percent were defendants of color while 36.6 percent were White/Non-Hispanic³. A quarter of the sample was 21 years of age or younger while the rest of the sample were distributed relatively evenly across 4 age subgroups.

	N	%
Gender		
Male	408	71.8
Female	160	28.2
Race		
White	208	36.6
Persons of Color	360	63.4
Age		
18-21	147	25.9
22-25	97	17.1

² One of the goals of creating a risk assessment is to ensure that it does not increase disproportionate minority contact at later stages. To reduce this risk, instruments should be constructed on the full population and then the developer should take steps to ensure that an individual's race or gender is not impacting the predictive validity of the tool. Where necessary, developers will create different cut points or even different predictors for people of color or female defendants to ensure that the tool is not over classifying non-white groups.

³ The Persons of Color include any defendant who was identified as African-American, Hispanic, Native American or Asian.

26-32	114	20.1
33-45	121	21.3
46 and older	89	15.7

Validation Results for the Virginia Pretrial Risk Assessment Instrument (VPRAI)

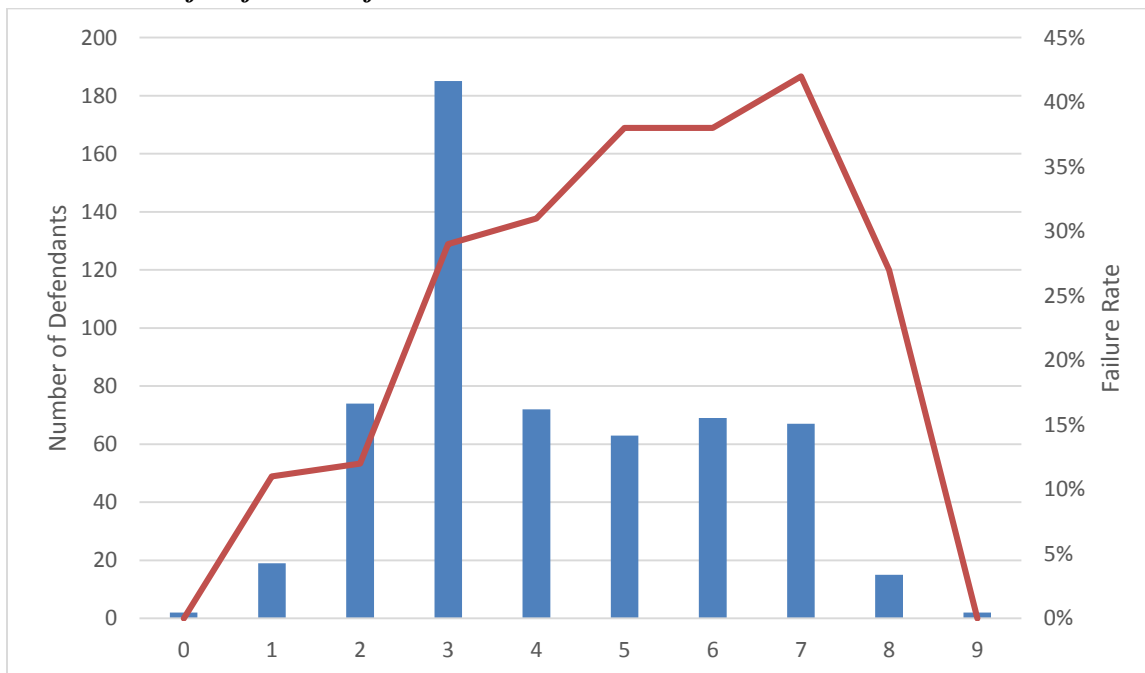
The first step in this study was to examine the validity of the Virginia Pretrial Risk Assessment Instrument (VPRAI) for the Riverside population. Chart 1 provides a visual of the sample by overall risk score as well as failure rates for each individual score.

Risk Score and Pretrial Failure

As shown below, 185 defendants scored a 3 on the VPRAI. Most other defendants fell between 2 and 7 points. The line graph represents failure rates by risk score. As the chart shows, the VPRAI is able to separate the failure rates by score.. The failure rates for risk scores of 8 and 9 should be interpreted with caution given the small numbers of defendants in that category.

Chart 1

Distribution of Defendants for the VPRAI Score and Failure Rate



$p \leq .0001$; $r = .186$; $AUC = .611$

Risk Level and Pretrial Failure

Table 3 provides the failure rates for each of the current risk categories. The overall ability of the VPRAI to predict failure is significant, but the current cut points do not provide substantive differences in the population. For example, the Level 1 failure rate is 9.5 percent, while the failure rate for Level 2 is only 2.7 percentage points higher. Similarly, the substantive difference between a Level 3 and Level 4 are negligible. This suggests that collapsing the VPRAI into three unique categories may provide more useful information.

Table 3
Validity of the VPRAI

	Failure	% Failure
0-1 Low Risk	2	9.5
2 Below Average Risk	9	12.2
3 Average Risk	53	28.6
4 Above Average Risk	22	30.6
5+ High Risk	82	38.0

$p \leq .01$; $r = .162$; $AUC = .609$ *

* The correlation and AUCs for the VPRAI change from the full score suggesting the actual raw score is slightly more predictive than the groups.

Adjusted Risk Level Categories

Given that the VPRAI was predictive of failure, the risk level cutoffs were recalculated to provide better utility in discerning between individuals who were more or less likely to fail. The adjusted cutoffs (Table 4) provide a clearer separation between low risk, moderate, and high risk. While the correlations and the AUC do not change significantly, the utility of 3 categories with substantively different failure rates make these cut points more practical for the program than the original cutoffs for the VPRAI.

Table 4
Re-Norming VPRAI Cutoffs

	Failure	% Failure
0 – 2 Low Risk	11	11.6
3 – 4 Moderate Risk	75	29.2
5+ High Risk	82	38.0

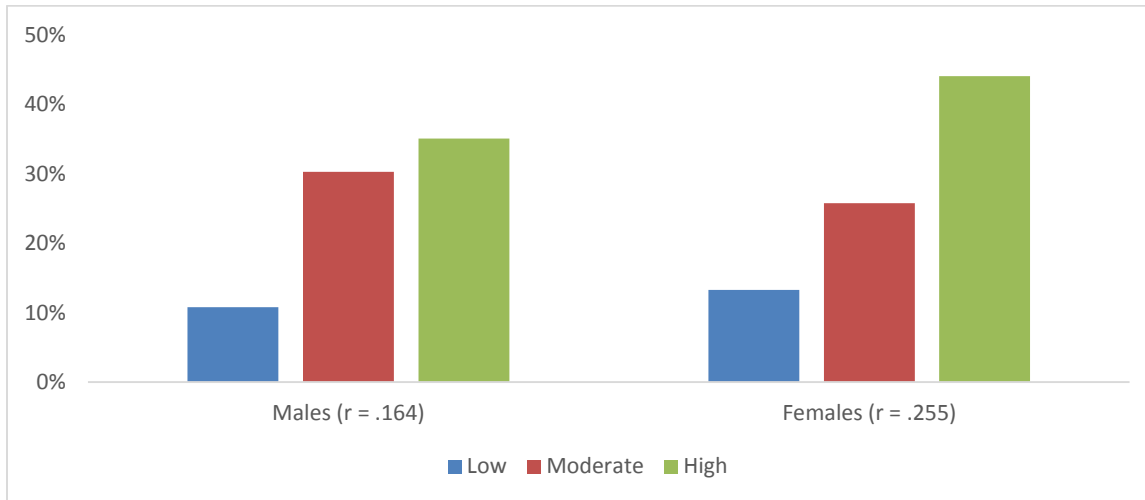
$p \leq .0001$; $r = .192$; $AUC = .609$

Examining the VPRAI by Subpopulations

While it is important to examine the effectiveness of the VPRAI for the total population, it is just as important to ensure that it is predictive for both males and females as well as people of color and Whites. To determine the validity by gender and race, the new cutoffs for the VPRAI were

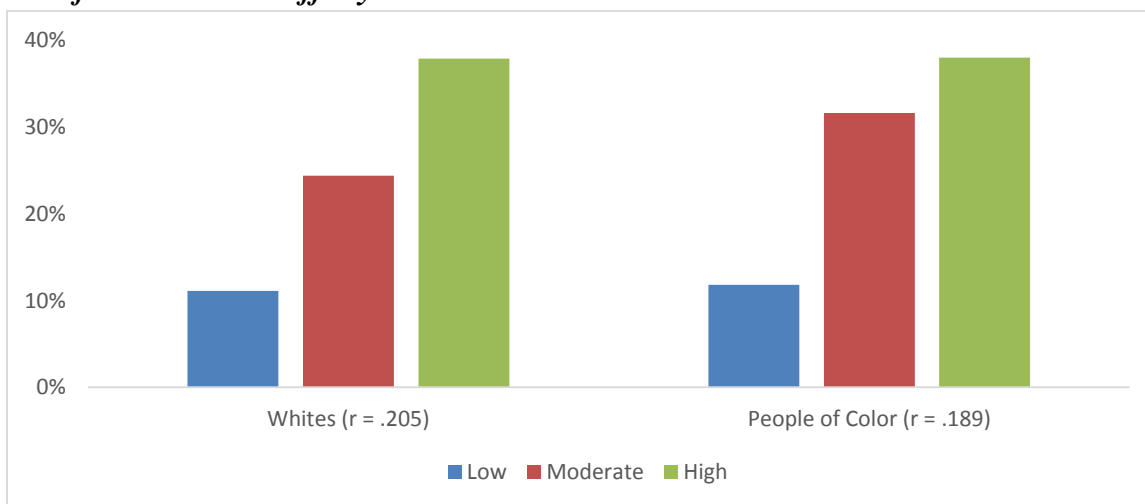
examined uniquely for males and females first. As noted in Chart 2, the VPRAI performed well for females, but did not do as well for males evidenced by the minimal difference between moderate and high risk males (5 percentage points).

Chart 2
Modified VPRAI Cutoffs by Gender



As for race, the VPRAI was predictive for both Whites and people of color. The low and high risk categories were almost identical across racial groups, but moderate risk defendants of color had a significantly higher failure rates than Whites. It should be noted that while predictive for all races, the VPRAI did not do especially well separating moderate and high risk defendants.

Chart 3
Modified VPRAI Cutoffs by Race



Beyond the VPRAI: A Revised Pretrial Risk Assessment Tool

In addition to examining the validity of the VPRAI, this study also set out to determine if adding any new items (or removing existing items from the VPRAI) could provide a more effective means to identify and separate defendants into risk categories. As noted in the previous section, the VPRAI was predictive of failure and with the new cutoffs was able to separate the population into 3 unique categories. This section sets forth to determine if the VPRAI can be improved by examining other predictors of risk. Table 4 provides a review of the bivariate relationships. The items that were significant predictors are shown in bold and marked (*).

Table 5
Predictors of new arrest/failure to appear

	% FTA/New Arrest	Sig level
Education Level		
HS Diploma/Greater	27.9	
Less than HS Diploma	32.5	
Current Charges		
2 or Fewer	29.5	
3 or More	29.9	
Current Charge Felony?		
No	27.6	
Yes	29.7	
*Pending Charges at time of New Offense		
No	26.8	p ≤ .01
Yes	40.0	
On Post-Sentence Supervision		
No	26.8	p ≤ .05
Yes	35.8	
1 or more Misd or Felony Conviction		
No	25.7	
Yes	32.8	
*Age at 1st Arrest		
21 or older	27.9	p ≤ .01
Under 21	43.3	
*Total # Prior Adult Arrests		
0	20.9	p ≤ .01
1 or more	32.9	

Table 5
Predictors of new arrest/failure to appear

		% FTA/New Arrest	Sig level
*2 or More Prior FTAs			
	No	26.3	p ≤ .05
	Yes	35.7	
*Number of Prior FTAs in Past 2 Years			
	None	25.3	p ≤ .01
	1	33.8	
	2 or more	40.5	
Prior Violent Convictions			
	1 or less	29.8	
	2 or more	24.0	
*Number of Previous Adult Convictions			
	1 or fewer	25.2	p ≤ .05
	2 or more	33.5	
Employed 2 years+			
	Yes	24.0	
	No	30.8	
*Current Residence 1+ Yr			
	Yes	25.2	p ≤ .05
	No	34.3	
*Substance Abuse History			
	No	25.9	p ≤ .05
	Yes	35.5	

While there are several methods to create a risk assessment based on the identified predictors, a Burgess Scale style is often favored over more technically complicated models that weight items. This is due to the simple scoring which allows for a straightforward, valid measure of risk which is easily interpretable by both defendants and staff.

While 9 items were statistically significant by themselves, when combined together some of these items were no longer significant because they were closely related to other items and therefore did not explain additional variation in predicting failures. Therefore, it was determined that the best combination of measures was a set of 5 items. Table 6 provides the items and the scoring associated with each item.

Table 6
Scoring for a Revised Pretrial Tool

Pending Charges at time of New Arrest	0 = No Pending Charges 1 = Yes, Pending Charges
2 or more failure to appears (FTA) in past 2 years	0 = 1 or fewer FTA in past 2 years 1 = 2 or more FTA in past 2 years
Substance abuse problem	0 = No substance abuse issue 1 = Has a substance abuse issue
Number of Previous Adult Convictions	0 = No previous adult convictions 1 = 1 or more previous adult convictions
1+ Year at Current Residence	0 = Yes 1 = No

Chart 3 provides the number of defendants by the revised risk instrument score and the subsequent failure rate associated with the individual scores. As illustrated in the chart, 77 defendants scored 0 and had a failure rate of 13 percent compared to a combined failure rate of 45 percent for those who scored 3 and 4.⁴ Furthermore, it is clear that there is a distinct difference between the scores, suggesting there are 3 naturally occurring categories, 0; 1 thru 2; 3 through 5.

⁴ While the failure rates continue to increase as the defendants' score increases, there were only 16 defendants who scored 5, making the results inconsistent. For example, with only 16 defendants it would only take 2 more defendants to fail to push the rate to 50 percent.

Chart 3

Number of Defendants by Risk Score and Failure rate

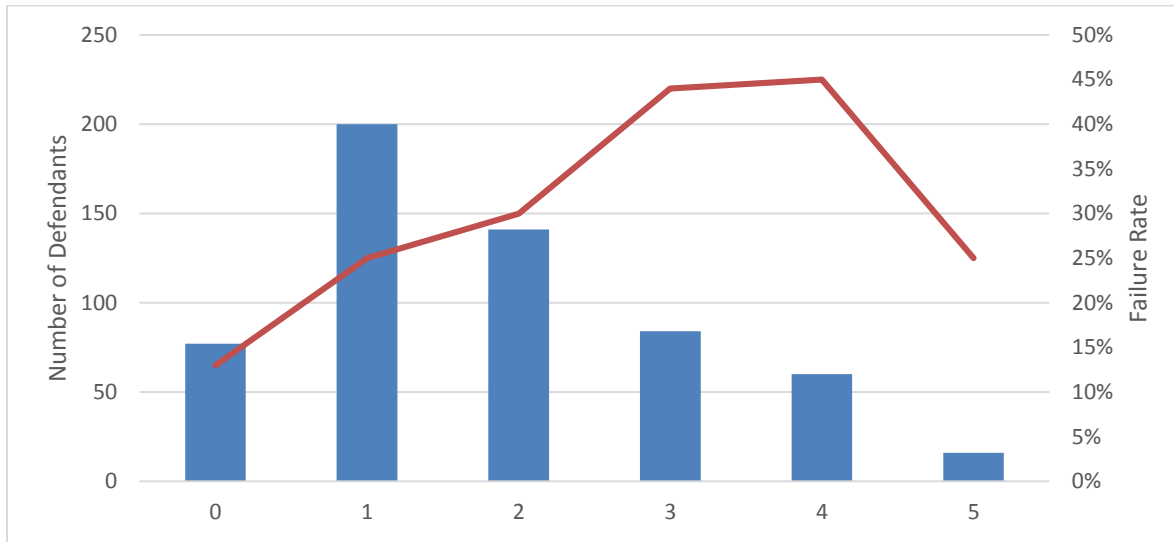


Table 8 provides the cutoffs for the suggested categories and the subsequent failure rates. The score of 0 resulted in a 13 percent failure rate and each subsequent level demonstrated a statistically significant increase in failures. Overall, the correlation of the revised assessment is .205 and the AUC is .614. The higher AUC provides evidence that the revised set of items is slightly more accurate than the original VPRAI (AUC = .611) or the VPRAI with modified cutoffs (AUC = .609) shown in Table 3 and Table 4.

Table 8
Cutoffs for New Items

	N	Failure	% Failure
0 – Low Risk	77	10	13.0
1 to 2 – Moderate Risk	331	90	27.2
3 to 5 – High Risk	160	68	42.5

p ≤ .0001; r = .205; AUC = .614

One of the important steps in developing a pretrial tool is to examine the effectiveness of such a tool to predict for subpopulations of the larger group. In this case, we are most interested in the revised pretrial tool’s ability to predict failure by gender and race to ensure that women and people of color are not over classified in higher risk categories than for the other populations. An example of over classification of a subgroup would be if the tool tended to classify a significantly higher percentage of women as high risk but in actuality the failure rate was significantly lower than men at a similar risk level. The next two charts demonstrate how a revised pretrial tool predicts failure by gender and race. Chart 5 shows a breakdown of failure

rates by gender.⁵ As indicated in the chart, the scale is predictive for both males and females, with almost identical failure rates between the genders.

Chart 5

Failure rates by risk level for males and females

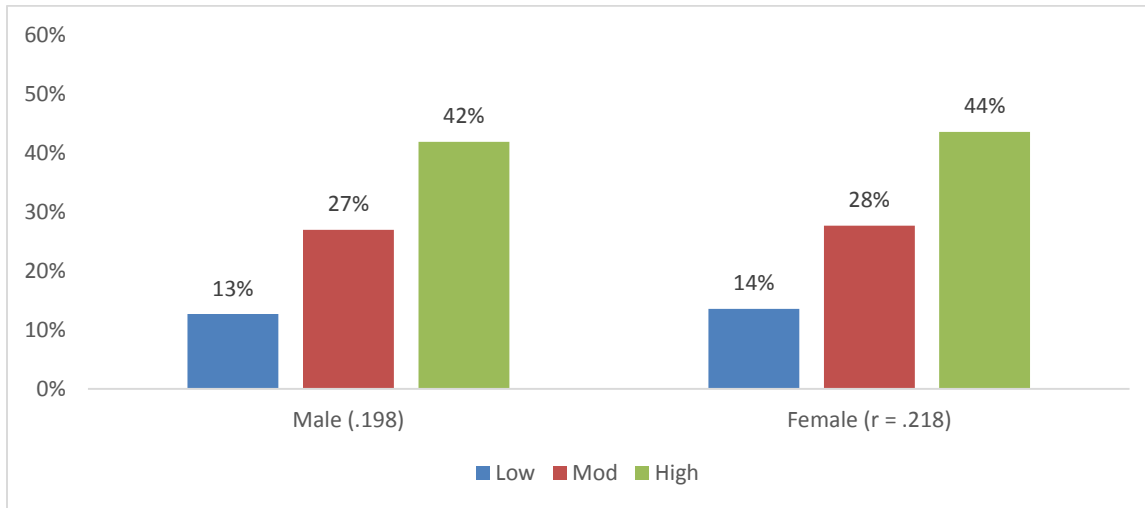
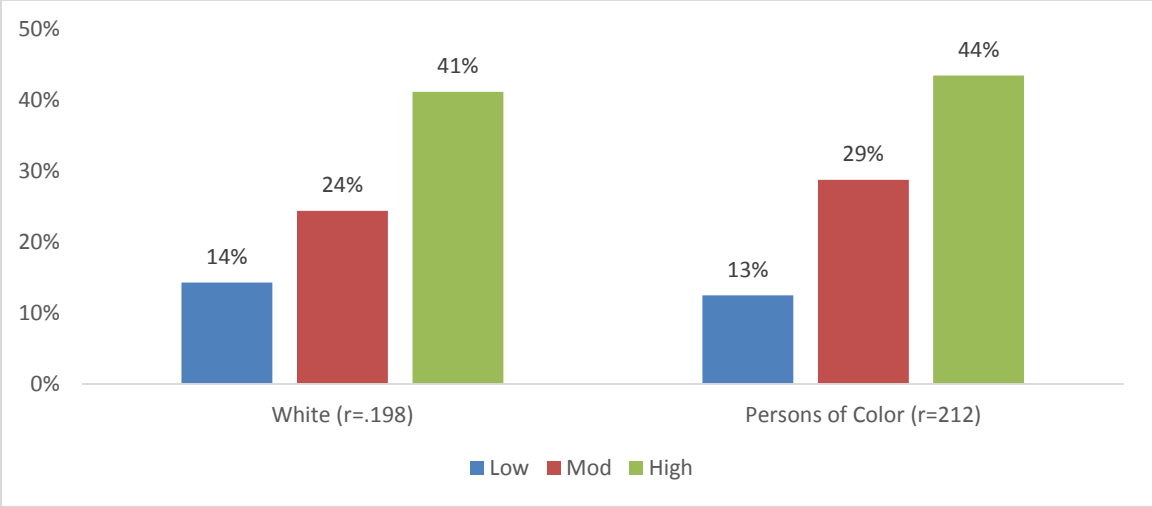


Chart 6 examines failure rates by race. The chart suggests that the items predict well for whites as well as persons of color, as indicated by the differentiation in failure rates across risk levels, though failure rates are higher in the moderate risk level for defendants of color compared to whites. This suggests that while the base rates of pretrial failure are higher for persons of color, the instrument demonstrated predictive validity for both whites and persons of color. The correlations for whites and persons of color are similar, suggesting that the tools identify defendants between risk levels equally well across race. Moreover, examining the failure rates across race/ethnicity, it is evident that the tool produces similar recidivism rates for Whites and Persons of Color.

Chart 6

Failure Rates by Risk Level for Whites and Persons of Color

⁵ It should be noted that there was a small sample of females in this study and therefore these results should be taken in context. It is recommended that the program continue to collect a subsample of data for further validation.



Implementation of the Modified Pretrial Tool

While it is important to adopt a validated pretrial tool, it is just as important to ensure that the tool that is selected be implemented appropriately. Riverside Probation and the PSU should continue their implementation efforts and explore additional opportunities to decrease the failure to appear rate:

1. Given the small sample size, especially for females, it is recommended that Riverside continue to collect data and repeat the validation analysis. Staff should continue to use the current data collection tool with one specific change:
 - a. Previous research has found that employment is predictive of future failure. The scoring rule for employment that was being used to collect the data was to determine if the client was employed for at least 2 years prior. This item was not valid, but it may be explained due to low variation on the item (only 17% of the defendants were identified employed). It is recommended that Riverside create a new item that examines shorter periods of employment. For example, employed at time of arrest or how long employed at current job instead of employed for 2+ years.
2. The department should modify policies and procedures to reflect changes to the risk tool items and scoring, and should train all staff and stakeholders on the new tool. The department should continue to train any new staff prior to conducting the assessments on the purpose of the pretrial tool and the scoring elements.
3. While it does not appear to be an issue in these data, Riverside Probation and the Pretrial Services Unit should consider examining inter-rater reliability as they move forward and expand the use of the tool.⁶ Examining inter-rater reliability will help to ensure accurate and consistent scoring across staff. To examine inter-rater reliability, the agency could do one of the following (or both):
 - a. Use a paired system in which one person gathers all the information while another person is observing the interview. Have both parties score out the tool independently and compare the scores. If there are conflicting results, have the two parties discuss the scoring and come to a consensus. Repeat this process for all staff conducting assessments to ensure that the scoring is similar across all staff.
 - b. Use a vignette or video of an interview and have individuals score independently of each other. Once the scoring is complete, have them compare scores and discuss any differences.

⁶ For any risk assessment, interrater reliability is important for the assessment to be valid. Specifically, if the staff is scoring the tool differently from one another it is impossible to determine if the tool is valid because the responses are not reliable.

4. The department should continue to employ strategies for release associated with each level of risk. The following are offered as suggestions and should be modified where appropriate for the resources available to the RCPD:
 - a. Low risk should be eligible for OR program and/or minimal services. These defendants should be given a court date and reminders sent prior to his or her court date. In total, the low risk group committed no new offenses during the follow up period and 89 percent of the defendants appeared as scheduled.
 - b. Strategies for low-moderate and moderate risk defendants should be developed to reduce the number of failure to appear warrants as well. While the failure rate for this group was driven primarily by defendants failing to appear. It is recommended that the program invest in technological approaches that will assist in reminding defendants of upcoming court dates. This may include text messaging, cell phone apps, email correspondence as well as calls and community visits.
 - c. Moderate-high and high risk defendants should be reviewed on a case-by-case basis as appropriate for pre-release services. These individuals pose a significant risk to fail and if released, should be provided very intensive services to ensure that they attend their court hearings. This should include some aspect of intensive monitoring (GPS, house arrest, etc) and should be mandated to any treatment services identified as appropriate, and all efforts afforded to ensure that the barriers to attending future court dates be removed (transportation, notifications, etc.).
5. The department should continue to use established override procedures and monitor the frequency of overrides. There are generally two types of overrides:
 - a. First is an override where the staff person identifies something specific to the defendant's situation that would either suggest the person is less or more risky than the tool has identified. To address this, the program should establish a protocol for overrides of the assessment and monitor overrides to ensure that they do not occur too often (generally more than 15 percent of the time) or too infrequent (less than 5 percent).
 - b. The second type of override is an administrative override. While we know seriousness of the offense does not predict pretrial failures, it should be taken into account when determining release and supervision practices. Generally, agencies use the risk score in combination with the seriousness of the offense to address both.
6. Continue to use the Community of Practice group as a way to get periodic feedback from staff in regards to barriers to accurately collecting information, scoring, or interpretation. This will help the agency ensure that the data is accessible and work towards longer term strategies if there are barriers to collecting and accessing the data efficiently.

References

- Cohen, T. and Reaves, B. (2007). Pretrial release of felony defendants in state courts. Washington, DC: Bureau of Justice Statistics, US Department of Justice.
- Cooprider, K. (2009). Pretrial risk assessment and case classification: A case study. *Federal Probation*, 73(1).
- Levin, D. (2007). *Examining the Efficacy of Pretrial Release Conditions, Sanctions and Screening with the State Court Processing Statistics Data Series*. Washington, DC: Pretrial Justice Institute.
- Lowenkamp, C. (2008). The development and validation of a pretrial screening tool. *Federal Probation*, 72(2).
- Nuttall, C.P; Barnard, E.F; Fowles, A.J; Frost, A; Hammond, W.H; Mayhew, P; Pease, K; Tarling, R and Weatheritt, M.J. (1977) *Parole in England and Wales, Home Office Research Study No. 38*. London, Home Office.
- Schwalbe, C. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and Human Behavior*; 31: 449-462.
- Summers, C. and Willis, T. (2010). *Pretrial Risk Assessments: Research Summary*. Washington, DC: Bureau of Justice Assistance, U.S. Department of Justice.
- VanNostrand, M. and Rose, K. (2009). Pretrial Risk Assessment in Virginia. Report for the Virginia Department of Criminal Justice Services, Richmond.
- Winterfield, L., Coggeshall, M. and Harrell, A. (2003). *Development of an Empirically-Based Risk Assessment Instrument: Final Report*. Washington, DC: Urban Institute Justice Policy Center.