Pretrial Risk Assessment Tool Validation

PRETRIAL PILOT PROGRAM

SEPTEMBER 2022

JUDICIAL COUNCIL OF CALIFORNIA
OPERATIONS AND PROGRAMS DIVISION
CRIMINAL JUSTICE SERVICES

CONTENTS

IMPACT OF THE COVID-19 PANDEMIC ON THE PRETRIAL PILOT PROGRAM

The Budget Act of 2019 requires that Pretrial Pilot Program courts collaborate with local justice system partners to make data available to the Judicial Council as required to measure the outcomes of the pilots. Senate Bill 36 (Hertzberg; Stats. 2019, ch. 589) established tool validation and reporting requirements for pretrial services agencies using a pretrial risk assessment tool; these requirements are mandatory for all pilot projects.

Throughout much of period covered by this report, the United States experienced the COVID-19 global pandemic. On March 4, 2020, Governor Gavin Newsom declared a state of emergency to protect public health and safety, and formalized efforts by the California Department of Public Health, California Health and Human Services Agency, Governor's Office of Emergency Services, and other state agencies and departments to mitigate this public health crisis. On March 19, 2020, orders from the Governor and the California Department of Public Health directed all California residents to stay home except when performing essential jobs or shopping for necessities.

On March 27, 2020, the Governor issued an order that gave the Judicial Council of California and the Chief Justice authority to adopt emergency rules and take other necessary actions to respond to the COVID-19 health and safety crisis. The Judicial Council adopted various emergency measures to support courts in providing essential services while helping to safely reduce jail populations. These measures, together with policies adopted by individual courts in response to the crisis, have impacted the population eligible for participation in the Pretrial Pilot Program.

On April 6, 2020, the Judicial Council adopted a statewide emergency bail schedule that set presumptive bail at $0 for most misdemeanors and lower-level felonies, with specified exceptions, but retained court discretion in setting bail. The emergency rule was intended to safely reduce jail populations and protect justice system personnel and public health while promoting consistency in pretrial release and detention throughout the state. The Judicial Council repealed the emergency bail schedule rule effective June 20, 2020 but encouraged courts to adopt local emergency bail schedules with $0 bail or significantly reduced bail levels to meet their county's public health and safety conditions.

As a result of local criminal justice system policies and the emergency bail schedule, pilot courts observed significant reductions in booking rates and jail populations during this time. Under these temporary emergency policies, many individuals who would otherwise have been eligible for program participation were cited and released in the field or released on $0 bail upon booking without undergoing a risk assessment. Crime and arrest patterns were also likely affected by COVID-19 and shelter-in-place orders. Criminal case dispositions also slowed during this period.

Therefore, the population of program participants is very likely different than would be seen in the absence of the pandemic, both in terms of reduced numbers and composition.

EXECUTIVE SUMMARY

This report fulfills the legislative mandates of the Budget Act of 2019 (Assem. Bill 74; Stats. 2019, ch. 23), and Senate Bill 36 (Stats. 2019, ch. 589). SB 36 requires each pretrial services agency that uses a pretrial risk assessment tool, including the Pretrial Pilot Projects, to validate the risk assessment tool used by the program by July 1, 2021, and on a regular basis thereafter, and to make specified information regarding the tool, including validation studies, publicly available. The Judicial Council is required to maintain a list of pretrial services agencies that have satisfied those validation requirements and complied with those transparency requirements. AB 74 also provided funding to the Judicial Council to assist the pretrial pilot courts in validating their risk assessment tools.

In response to the requirements of AB 74 and SB 36, the Judicial Council of California conducted the following validation studies for four pretrial risk assessment tools. The period for this validation extends from October 1, 2019, to December 31, 2021, and includes a diverse sample of counties in California. Among the counties, population size ranges from less than 10,000 to over 10 million; geographic regions range north to south as well as inland and coastal. Demographically, the counties represent both majority Hispanic and Non-Hispanic White populations.

The Judicial Council conducted pretrial risk assessment validation studies for:

- Ohio Risk Assessment System Pretrial Assessment Tool (ORAS-PAT), developed by the University of Cincinnati, Center for Criminal Justice Research—used by the Pretrial Pilot Projects in Modoc, Napa, Nevada/Sierra, Ventura, and Yuba Counties.
- Public Safety Assessment (PSA), developed by Arnold Ventures—used by the Pretrial Pilot Projects in Calaveras, Los Angeles, Sacramento, Sonoma, and Tuolumne Counties.
- Virginia Pretrial Risk Assessment Instrument (VPRAI), developed by the Virginia Department of Criminal Justice Services—used by the Pretrial Pilot Project in San Joaquin County.
- Variations of the VPRAI (VPRAI-R and VPRAI-O)—used by the Pretrial Pilot Projects in Alameda, Kings, and San Mateo Counties.

This report includes the following validation studies:

- Overall validations for the ORAS, PSA, VPRAI, and VPRAI-R tools, and a study of predictive validity of the tools by race/ethnicity and gender.

If larger sample sizes become available in the future, the Judicial Council will conduct a validation study for the VPRAI-O risk assessment tool.

The risk scores presented in this report are calculated using a scoring scheme designed by tool developers. The tools account for aspects of an individual's criminal history, current criminal offense, history of failures to appear in court, age, and other factors depending on the tool. (See Appendix A for the factors and weights specific to each tool.) Gender and race are not used in any of the tools to calculate risk scores.

In conducting the validations, the Judicial Council used the Area Under the Curve (AUC) and logistic regression to examine the tools' accuracy and reliability. The AUC value is a single number that represents the ability of the tool to differentiate between individuals who are lower or higher risk across the range of the tool. The AUC is calculated for each tool overall and, when the sample size is sufficient, separately for each gender and race/ethnicity group to examine whether the ability of the tool to differentiate individuals by risk differs by gender or race/ethnicity. Logistic regression is used to test whether risk scores statistically significantly predict the likelihood of each outcome of interest (failure to appear; new arrest; new filing; new conviction; new violent arrest; and a composite measure of FTA or new arrest), and whether any differences in outcomes by risk level across gender or race/ethnicity are statistically significant. Statistical significance is a technical term used in analyses to indicate that it is very unlikely that a result or difference occurred by chance. Statistical significance does not necessarily indicate the size of the result or difference.

In the validation studies conducted by the Judicial Council, using a common metric for interpreting AUC values in criminal justice risk assessments, when the entire sample was considered, all tools had AUCs in the good-to-excellent ranges for all outcomes except new violent arrest. AUCs for new violent arrests tended to fall in the fair range for most tools. Some differences in AUCs by gender and race were apparent when subgroups were compared, but there was no persistent pattern across tools.

Results of the regression analyses show that the association between risk score and all outcomes of interest was statistically significant. Race/ethnicity was found to be a statistically significant predictor of some outcomes for some race/ethnic group for all tools. Further research is needed to analyze the elements that may be driving the observed differences and whether there are data-driven modifications to the tools' risk factors or weights that can further improve the predictive power of the tools. Gender was also a significant predictor of some outcomes for all tools except the VPRAI-R. Generally, the tools tend to overpredict the risk of adverse outcomes for females relative to males.

This report solely analyzes risk scores and associated outcomes for individuals who were released pretrial. Individuals may have been released by the sheriff, by a judge, or on bail. This report does not look at judicial decisionmaking or judges' use of the tools.

We would like to thank the courts and their justice partners in the pilot counties for their participation in these validation studies.

INTRODUCTION

LEGISLATIVE MANDATE

This report fulfills the legislative mandates of the Budget Act of 2019 (Assem. Bill 74; Stats. 2019, ch. 23), and Senate Bill 36 (Stats. 2019, ch. 589). In AB 74, the Legislature directed the Judicial Council to administer two-year pretrial projects in the trial courts. The goals of the Pretrial Pilot Program, as set by the Legislature, are to:

- Increase the safe and efficient prearraignment and pretrial release of individuals booked into jail;
- Implement monitoring practices with the least restrictive interventions necessary to enhance public safety and return to court;
- Expand the use and validation of pretrial risk assessment tools that make their factors, weights, and studies publicly available; and
- Assess any disparate impact or bias that may result from the implementation of these programs.

Sixteen Pretrial Pilot Projects (17 courts) were selected to participate in the program.[1]

SB 36 requires each pretrial services agency that uses a pretrial risk assessment tool, including the Pretrial Pilot Projects, to validate the risk assessment tool used by the program by July 1, 2021, and on a regular basis thereafter, and to make specified information regarding the tool, including validation studies, publicly available. The Judicial Council is required to maintain a list of pretrial services agencies that have satisfied those validation requirements and complied with those transparency requirements. The Judicial Council is also required to publish a report on the judicial branch's public website with data related to outcomes and potential biases in pretrial release.

AB 74 provided funding to the Judicial Council "for costs associated with implementing and evaluating these programs, including, but not limited to: [¶] ….(e) Assisting the pilot courts in validating their risk assessment tools." This report, in accordance with AB 74 and SB 36, provides information on the validation of the pretrial risk assessment tools used by the 17 Pretrial Pilot courts.

Pretrial risk assessment tools use actuarial algorithms to assess the likelihood that a person who has been arrested for an offense will fail to appear in court as required or will commit a new offense during the pretrial period. The pretrial risk assessment tools used by the 16 Pretrial Pilot Projects are:

- ORAS (Ohio Risk Assessment System; developed by the University of Cincinnati, Center for Criminal Justice Research);
- PSA (Public Safety Assessment, developed by Arnold Ventures);

[1] The pilot counties are Alameda, Calaveras, Kings, Los Angeles, Modoc, Napa, Nevada, Sacramento, San Joaquin, San Mateo, Santa Barbara, Sierra, Sonoma, Tulare, Tuolumne, Ventura, and Yuba Counties. Nevada and Sierra Counties participated as a consortium.

- VPRAI (the Virginia Pretrial Risk Assessment Instrument, developed by the Virginia Department of Criminal Justice Services); and
- VPRAI-R and VPRAI-O (variations of the VPRAI).

SB 36 requires pretrial risk assessment tools to be validated. SB 36 defines "validate" as follows:

> "Validate" means using scientifically accepted methods to measure both of the following:
> (A) The accuracy and reliability of the risk assessment tool in assessing (i) the risk that an assessed person will fail to appear in court as required and (ii) the risk to public safety due to the commission of a new criminal offense if the person is released before the adjudication of the current criminal offense for which they have been charged.
> (B) Any disparate effect or bias in the risk assessment tool based on gender, race, or ethnicity.[2]

VALIDATION METHODS

The following methodology has been used to validate each of the pretrial risk assessment tools used by Pretrial Pilot Projects for which data were sufficient.

Descriptive statistics are presented, exploring basic features of the data such as demographics, and showing the overall distributions of arrest offenses and adverse outcomes. The distributions of risk scores are shown in groupings of risk level defined by each tool developer.

A receiver operating characteristic (ROC) curve model has been used to provide the area under the curve (AUC) statistic for each outcome of interest. The outcomes of interest are:

- Failure to appear (FTA);
- New arrest;
- New filing;
- New conviction;
- New violent arrest; and
- FTA or new arrest (composite measure).

The AUC value is a single number that represents the ability of the tool to differentiate between individuals at lower or higher risk across the range of the tool. The AUC is calculated for each outcome overall and separately for each gender and race/ethnicity group to examine whether the ability of the tool to differentiate individuals by risk differs by gender or race/ethnicity.

[2] Sen. Bill 36, § 1320.35(b)(4).

For criminal justice risk assessments, a common metric for evaluating AUC values is derived from Desmarais and Singh (2013),[3] who defined AUC values less than 0.55 as poor, 0.55-0.63 as fair, 0.64-0.70 as good, and 0.71-1.00 as excellent.

The observed rate of adverse outcomes at each score is presented. The pattern of these rates is an indicator of the accuracy of the tool, showing whether risk scores predict monotonic increasing failure rates for each outcome of interest.

Logistic regression is used to test whether the risk scores statistically significantly predict the likelihood of each outcome of interest and whether any differences in outcomes by risk level across gender or race/ethnicity are statistically significant. Statistical significance is a technical term used in analyses, to indicate that it is very unlikely that a result or difference occurred by chance. Statistical significance does not necessarily specify the size of the result or difference.

To measure any predictive bias in the tools, fitted curves of the rates of adverse outcomes at each score are shown separately by gender and race/ethnicity groups. Logistic regression has been used to test whether the likelihood of each outcome of interest by risk level differs statistically significantly across gender or race/ethnicity groups.

This report solely analyzes risk scores and associated outcomes for individuals who were released from custody pretrial. Individuals may have been released in a variety of ways by the sheriff or judge, including on bail. This report does not look at judicial decisionmaking or judges' use of the risk assessment tool.

Further research is needed to analyze the elements that may be driving the observed differences and whether there are data-driven modifications to the tools' risk factors or weights that can further improve the predictive power of the tool.

DEFINITIONS

- **Pretrial period** starts at the booking of an individual at the jail and ends at the resolution of all cases associated with that booking.
- **Failure to appear** (FTA) is measured using court records documenting the issuance of a bench warrant for FTA during the pretrial period.
- **New arrest** is any new arrest during the pretrial period reported to the California Department of Justice (CA DOJ) or a new booking within county recorded by the jail.[4]

[3] Sarah L. Desmarais and Jay P. Singh, *Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States* (Lexington, KY: Council of State Governments, 2013).
[4] New criminal offenses are defined in four ways to capture different outcomes of interest. All new criminal offense indicators are measured using data from the California Department of Justice (CA DOJ).

- **New filing** is any new arrest during the pretrial period that results in charges filed with the court and reported to the CA DOJ.[5]
- **New conviction** is any new arrest during the pretrial period that results in a conviction reported to the DOJ during the data collection period.[6]
- **New violent arrest** is any new arrest during the pretrial period for an offense on the Pretrial Pilot consensus PSA Violent Offense List, which includes felonies and misdemeanors of a violent nature. For the full list of offenses, see Appendix B.
- **FTA or new arrest** is a combined measure indicating an occurrence of an FTA, a new arrest, or both. This measure is shown for the ORAS, VPRAI, and VPRAI-R, which were designed to predict overall "pretrial failure." This measure is not shown for the PSA because it was designed to predict outcomes separately.

VALIDATION SAMPLE SIZES

For purposes of this report, general validation results are shown when the sample size was greater than 200. For analyses of predictive bias by race/ethnicity and gender, subgroup results are shown when the overall sample was at least 1,000 and each subgroup size was greater than 200. Sample sizes smaller than these may not produce reliable results.

DATA DESCRIPTION AND LIMITATIONS

The data frame for the pretrial risk assessment tool validations was created using data from the court and two agencies in each of the Pretrial Pilot Program counties, as well as statewide data from the California Department of Justice. Although the number of assessed bookings during this period totaled 270,699, the evaluation data set used in this validation tracks the records of 50,875 bookings with associated pretrial risk assessments and completed pretrial periods. Assessed bookings without completed pretrial periods or for which the individual was not released pretrial are not included in the evaluation data set.

The risk scores presented in this report are calculated using a scoring scheme designed by the tool developers. The tools take into account aspects of an individual's criminal history, current criminal offense, history of failures to appear in court, age, and other factors (see Appendix A). Gender and race are not used to calculate risk scores.

DATA SOURCES

- **Jail booking data**: County sheriffs' offices provided information on all individuals booked into local county jails, including booking dates, charges, and releases.

[5] CA DOJ records on arrests are likely more complete than CA DOJ records on court filings and dispositions. Court reporting to the CA DOJ is incomplete.

[6] Due to the short time frame of the data collection period and delays in court reporting to the DOJ, new convictions may not be a complete measure of all arrests during the pretrial period that resulted in a conviction.

- **Probation data**: County probation departments performed pretrial assessment services and provided pretrial risk assessment information, including assessment dates, scores, and recommendations for those assessed.
- **Court case data**: Superior courts provided court case information, including pretrial disposition dates and the issuance of warrants for failures to appear for individuals people with felony or misdemeanor criminal filings.
- **California Department of Justice Data:** CA DOJ provided arrest and disposition data, including out-of-county filings, for booked defendants.

DATE RANGE

The time frame for this validation extends from October 1, 2019, to December 31, 2021, for most counties.[7] October 1, 2019, marks the beginning of the Pretrial Pilot Program grant period.

Table 1. Assessment Date Ranges, By Tool

Tool Name	Earliest Assessment Date	Latest Assessment Date
ORAS	2019-10-01	2022-02-23
PSA	2018-07-03	2022-07-13
VPRAI	2018-01-01	2022-01-05
VPRAI-R	2020-01-30	2022-02-03

DATA LINKING AND FILTERING

Data were viewed based on a data sharing agreement and data views were joined and standardized to create a validation data frame of bookings with associated pretrial risk assessment information, relevant court case information, and outcomes during the pretrial period. Only a small subset of the assessments conducted were used in the validation dataset. In some instances, not all data could be matched across agencies. The only bookings included in the validation data set were those for which the individual was released pretrial, and there was a final disposition associated with the booking because (1) outcomes during the pretrial period were a primary interest of this analysis, and (2) so that the full pretrial period could be observed. This report refers to each booking linked with an associated assessment and completed pretrial period as a "pretrial observation."

Table 2 and Table 3 show the number of assessments for each tool and county at each stage of filtration, and the type of validation that will be presented based on the sample size. It is anticipated that

[7] See Appendix C for date range for each county.

validation of all tools for all counties will be completed when sample sizes reach the thresholds described under Validation Sample Sizes, above.

The number of assessments performed during the evaluation period ranged from a low of 1,173 assessments for the VPRAI-O (Kings County) to a high of 254,872 assessments for the PSA (Calaveras, Los Angeles, Sacramento, Sonoma, and Tuolumne Counties). The next column represents assessments linked to unique jail bookings, ranging from a low of 1,269 for the VPRAI-O to a high of 230,350 for the PSA. The next column shows the number of bookings with associated pretrial risk assessments that have a final disposition.

Linking bookings with pretrial risk assessments and selecting only cases with a final disposition lowers the sample to a range of 300 to 113,984 observations. Because of the limited time period for the evaluation, this drop in observations is expected.[8] The next column shows the evaluation sample of bookings with associated pretrial risk assessments that have a final disposition, and in which the defendant was released pretrial. The evaluation samples range from 13 to 37,949 pretrial observations.

A large portion of the time frame for the evaluation data set overlapped with the COVID-19 pandemic, which likely had large impacts on crime, policing and booking practices, and the ability of courts to process cases, likely lowering the number of cases with a final disposition during this time frame. Because the sample size for the VPRAI-O was smaller than the designated minimum sample size of 200 for general validation, it was not possible to validate the VPRAI-O risk assessment tool in the current validation study.

Table 2. Counts of All Assessments at Each Stage of Filtration Sample, by Tool

Tool Name	Assessments	Assessed Bookings	Pretrial Complete	Validation Dataset	Validation Type
ORAS	10,886	9,407	5,566	3,256	General + Bias
PSA	254,872	230,350	113,984	37,949	General + Bias
VPRAI	13,150	12,886	10,499	5,681	General + Bias
VPRAI-R	18,101	18,056	8,550	3,989	General + Bias
VPRAI-O	1,173	1,269	300	13	Sample Too Small

[8] There is a smaller relative drop in observations from stage to stage with the VPRAI tool because the time frame of the VPRAI assessments is broader. (See Table 1 for date ranges for each tool.)

Table 3. Counts of All Assessments at Each Stage of Filtration for Evaluation Sample, by County

Tool Name	County	Assessments	Assessed Bookings	Pretrial Complete	Validation Dataset	Validation Type
ORAS	Ventura	7,062	6,786	4,327	2,835	General + Bias
PSA	Los Angeles	169,149	170,298	76,411	24,982	General + Bias
PSA	Sacramento	69,863	44,450	30,032	10,132	General + Bias
PSA	Sonoma	5,236	4,856	3,320	1,546	General + Bias
PSA	Tulare	10,208	10,325	3,922	1,146	General + Bias
VPRAI	San Joaquin	13,150	12,886	10,499	5,681	General + Bias
VPRAI-R	Alameda	12,464	12,444	4,745	2,439	General + Bias
VPRAI-R	San Mateo	5,637	5,612	3,805	1,550	General + Bias
ORAS	Modoc	284	273	138	57	Sample Too Small
ORAS	Napa	2,576	1,390	517	155	Sample Too Small
ORAS	Nevada/Sierra	482	266	112	43	Sample Too Small
ORAS	Yuba	482	692	472	166	Sample Too Small
PSA	Calaveras	353	353	262	115	Sample Too Small
PSA	Tuolumne	63	68	37	28	Sample Too Small
VPRAI-O	Kings	1,173	1,269	300	13	Sample Too Small

DESCRIPTIVE STATISTICS

PRETRIAL TOOLS

This report addresses the validation of four pretrial risk assessment tools: Ohio Risk Assessment System–Pretrial Assessment Tool (ORAS-PAT), Public Safety Assessment (PSA), Virginia Pretrial Risk Assessment Instrument (VPRAI), and Virginia Pretrial Risk Assessment Instrument–Revised (VPRAI-R). The Pretrial Pilot Program counties using each of these tools and included in the evaluation data set are listed in Table 4.

Table 4. Counties Contributing Assessment Data for Each Assessment Tool

Tool Name	County	Pretrial Observations
ORAS	Modoc, Napa, Nevada, Ventura, Yuba	3,256
PSA	Calaveras, Los Angeles, Sacramento, Sonoma, Tulare, Tuolumne	37,949
VPRAI	San Joaquin	5,681
VPRAI-R	Alameda, San Mateo	3,989
All	**All**	**50,875**

DEMOGRAPHICS

The pretrial programs evaluated in this validation come from a diverse sample of counties in California. Among the counties, population sizes range from less than 10,000 to over 10 million, and geographic regions range north to south as well as inland and coastal. Demographically, the counties represent both majority Hispanic and non-Hispanic white populations.

Additionally, there are broad differences in the racial and ethnic makeup of the assessed populations in each county. For each pretrial risk assessment tool used, Table 5 provides the number of assessments in the evaluation data set, the racial/ethnic and gender makeup, and the median age. The proportions of each race and ethnicity vary widely across tools (6–33% Black, 36–54% Hispanic, and 21–37% white), gender[9] proportions vary moderately (16–22% female), and median age varies slightly (34–36 years). This pattern of variation across counties in criminal justice–involved populations is typical.[10]

Table 5. Demographic Profile of Evaluation Data Set, by Tool

		Race/Ethnicity (%)				Gender (%)		
Tool Name	Total	Black	White	Hispanic	Other	Male	Female	Median Age
ORAS	3,256	6	37	54	3	78	22	35
PSA	37,949	25	23	47	5	81	17	34
VPRAI	5,681	22	28	40	10	80	20	35
VPRAI-R	3,989	33	21	36	11	84	16	36
Total	**50,875**	**24**	**24**	**45**	**6**	**81**	**17**	**34**

ARREST OFFENSES

The arrest offenses leading to the bookings in the evaluation data set varied across counties. Felony arrests represented the majority of bookings (66–79%); misdemeanor arrests were a smaller share (19–32%).[11] Violent offenses[12] represented 24–31% of bookings in the data set, property offenses were 17–24%, and drug offenses were 13–21% of bookings in the data set. Driving under the influence (DUI) offenses ranged from 5–16% of bookings, and domestic violence (DV) offenses made up 16–32% of bookings in the evaluation data set.

[9] Nonbinary, other, and unknown genders represented less than 0.1% of the bookings in the evaluation data set.
[10] See www.ppic.org/publication/californias-county-jails/; www.ppic.org/interactive/interactive-arrests-in-california/.
[11] The sum of felony and misdemeanor offense may not add up to 100% because infractions were not included in the table.
[12] Violent offenses, as defined by the pilot consensus PSA Violent Offense List (see Appendix B). These offenses include both felonies and misdemeanors that are violent in nature.

Table 6. Distribution of Arrest Offense Type, by Tool (numbers shown in percentages)

County	Felony	Misdemeanor	Violent	Property	Drug	DUI	DV
ORAS	79	19	31	17	19	5	32
PSA	66	32	24	21	21	16	16
VPRAI	68	22	27	20	13	10	26
VPRAI-R	74	26	25	24	20	5	19

ADVERSE OUTCOMES

Several different adverse outcomes are measured during the pretrial period from pretrial release to disposition. Failure to Appear (FTA), measured as bench warrants issued for FTA during the pretrial period, ranged from 22.0–32.6% of pretrial observations. New arrests during the pretrial period ranged from 41.5–54.3% of pretrial observations. New arrests during the pretrial period resulting in filed charges were recorded for 15.6–31.1% of pretrial observations, and new arrests during the pretrial period resulting in convictions were recorded for 12.3–23.8% of pretrial observations.[13] New violent arrests[14]—including felony and misdemeanor arrests for offenses of a violent nature—were recorded during the pretrial period for 8.8–19.6% of pretrial observations.

Table 7. Rates of Pretrial Misconduct, by Tool

Tool Name	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest
ORAS	23.5	41.8	15.6	12.3	8.8
PSA	22.0	41.5	20.8	15.6	11.7
VPRAI	30.4	49.1	31.1	23.8	19.6
VPRAI-R	32.6	54.3	25.8	14.7	17.1

CONDITIONS OF MONITORING/SUPERVISION

Data on supervision conditions were collected from county probation departments. However, data were of variable quality and were therefore not used in the analyses. Supervision conditions may have affected outcomes and may have been applied differentially according to risk score, which could confound results. Further research is needed to determine the impact of supervision conditions and to separate the efficacy of the tools from the efficacy of supervision conditions.

[13] New arrest, new filing, and new conviction data are measured using CA DOJ data. New arrests and new violent arrests are reported to the CA DOJ from arresting agencies, whereas new filings and new convictions are reported to the CA DOJ from courts. The DOJ may have incomplete records of filings and convictions from the courts because of difficulties or delays in reporting, and not all new arrests during the pretrial period may have been resolved during the data collection period.

[14] New violent arrests are defined by the PSA Violent Offense List (see Appendix B).

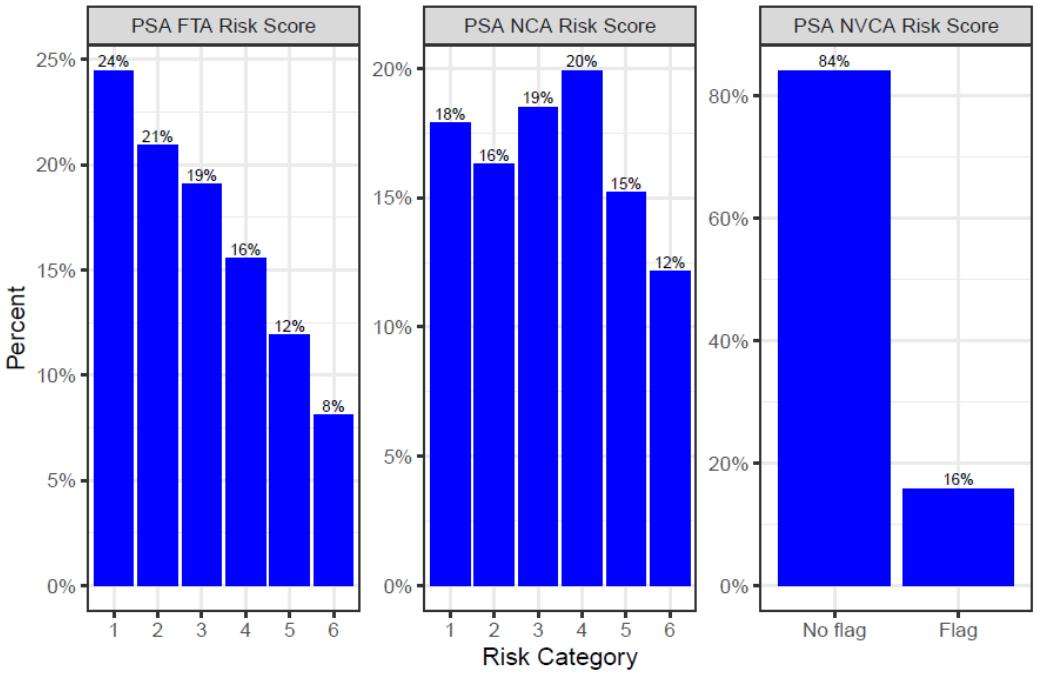
PUBLIC SAFETY ASSESSMENT (PSA) VALIDATION

GENERAL VALIDATION

Figure 1 shows the distribution of risk categories for individuals in the evaluation data frame assessed with the PSA tool—used by the Pretrial Pilot Projects in Calaveras, Los Angeles, Sacramento, Sonoma, Tulare, and Tuolumne Counties—for each PSA subscale. The PSA Failure to Appear (FTA) subscale was designed to predict the risk of failure to appear in court, the PSA New Criminal Activity (NCA) subscale was designed to predict the risk of a new arrest, and the PSA New Violent Criminal Activity (NVCA) flag was designed to predict the risk of a new arrest for a violent crime. As determined by the tool developers, the FTA and NCA subscales are each divided into 6 risk levels, with 1 representing the lowest risk and 6 the highest, and the NVCA subscale is divided into a binary flag, such that a flag represents higher risk of new violent crime, and no flag represents lower risk of new violent crime.[15]

Lower scores were more common for the FTA subscale. Of those assessed, 64% scored 1, 2, or 3 on the FTA subscale. The NCA subscale was less concentrated at the low end of the scale, with 53 percent receiving a score of 1, 2, or 3. Sixteen percent of assessed individuals received an NVCA flag. The distribution of all assessed individuals may differ from the distribution in the evaluation data frame because the evaluation data frame includes only released individuals with concluded pretrial periods. Table 8 shows the counts associated with the risk distributions in Figure 1.

Figure 1. Distribution of PSA FTA, NCA, and NVCA Risk Scores



[15] The NVCA subscale is scaled to a 1–6 scale. Scores 1–3 are categorized as not flagged, and scores 4–6 are categorized as flagged.

Table 8. Counts of Individuals by PSA FTA, NCA, and NVCA Risk Scores

PSA FTA Risk Score	Total
1	9,283
2	7,939
3	7,234
4	5,895
5	4,520
6	3,078

PSA NCA Risk Score	Total
1	6,792
2	6,191
3	7,025
4	7,558
5	5,775
6	4,608

PSA NVCA Risk Score	Total
No flag	31,938
Flag	6,011

Table 9 shows the AUC values for the PSA subscales for each outcome of interest. The AUC value is a single number that represents the ability of the tool to discriminate between individuals who are lower or higher risk across the range of the tool. For criminal justice risk assessments, a common metric for evaluating AUC values is derived from Desmarais and Singh (2013),[16] who defined AUC values less than 0.55 as poor, 0.55–0.63 as fair, 0.64–0.70 as good, and 0.71–1.00 as excellent. By these definitions, the AUC values for the PSA are (1) excellent for new arrest; (2) good for FTA, new filing, and new conviction; and (3) fair for new violent arrest.

The 95% confidence interval (CI) is also shown. It represents the range of AUC estimates that the true AUC value is 95% likely to fall between. A smaller range indicates that, given sample size and data pattern, the AUC can be estimated with greater precision. None of the 95% confidence intervals fall below the fair range.

[16] See note 3.

Table 9. AUC Values by Outcomes of Interest and PSA Subscale Risk Scores

Risk Score	Outcome	AUC	CI (95%)
PSA FTA	FTA	0.683	0.677-0.689
PSA NCA	New Arrest	0.712	0.707-0.717
PSA NCA	New Filing	0.680	0.674-0.686
PSA NCA	New Conviction	0.676	0.669-0.683
PSA NVCA	New Violent Arrest	0.580	0.573-0.587

N = 37949

Figure 2 shows the rate of various adverse outcomes during the pretrial period at each risk category of the PSA, using each of the PSA subscales for the relevant outcomes. The PSA-FTA risk scale is used for the outcome of FTA. The PSA-NCA risk scale is used for the outcomes of new arrest, new filing, and new conviction. The PSA-NVCA risk flag is used for the outcome of new violent arrest. For each outcome of interest,[17] observed rates of the outcome increase as the assessed risk level increases. This pattern is consistent across all outcomes and risk levels.

Figure 2. PSA Outcomes by Risk Category

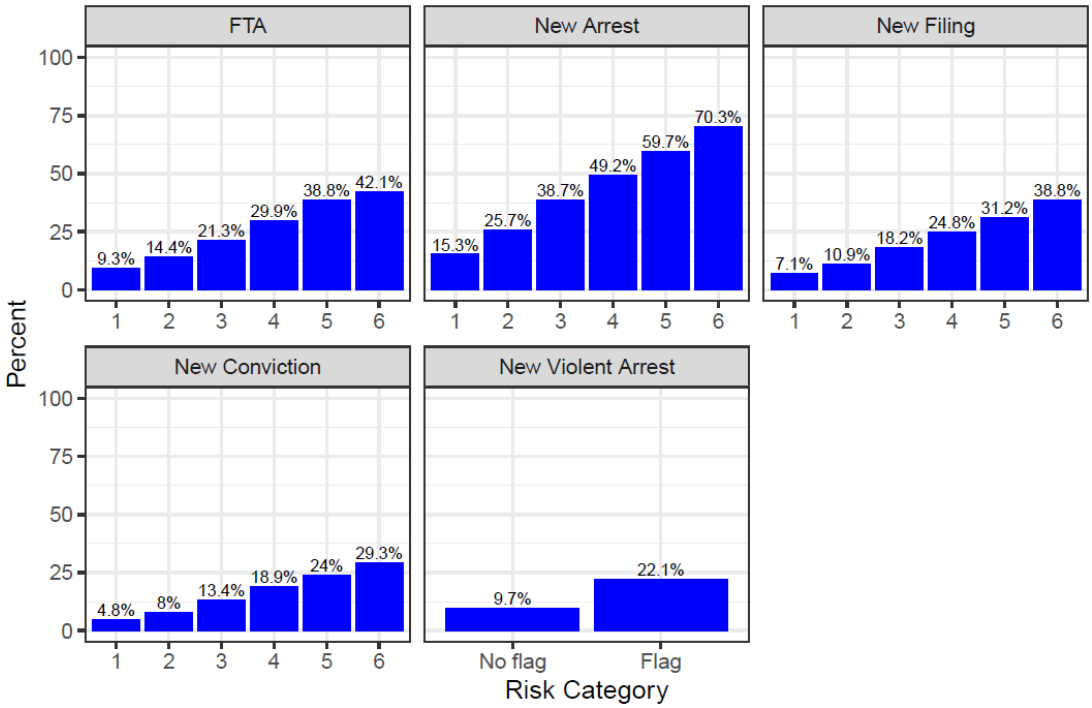


Table 10 shows the results from logistic regression models predicting each outcome of interest. The models control for the number of days the defendant spent released during the pretrial period. For each

[17] See the validation methodology section for definitions of each outcome of interest.

outcome of interest, the models show that the association between the relevant PSA risk score and the likelihood of the outcome during the pretrial period are statistically significant (p<0.001).

Table 10. Logistic Regression Models Testing the Likelihood of Outcomes of Interest by PSA FTA, NCA, and NVCA Risk Scores Controlling for Days Released

	Dependent variable:				
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest
	(1)	(2)	(3)	(4)	(5)
PSA FTA Risk Score	0.434***				
	(0.010)				
PSA NCA Risk Score		0.625***	0.481***	0.460***	
		(0.010)	(0.010)	(0.011)	
PSA NVCA Risk Score					1.172***
					(0.048)
Days Released	0.003***	0.005***	0.003***	0.003***	0.003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	−3.213***	−3.181***	−3.584***	−3.820***	−2.801***
	(0.047)	(0.046)	(0.052)	(0.057)	(0.039)
Observations	24,972	24,972	24,972	24,972	24,972
Log Likelihood	−11,580.530	−13,798.860	−11,769.410	−10,264.790	−8,621.007
Akaike Inf. Crit.	23,167.050	27,603.720	23,544.810	20,535.580	17,248.010
Note:					*p<0.05; **p<0.01; ***p<.001

ANALYSIS OF PREDICTIVE BIAS

RACE/ETHNICITY

Figure 3 shows the distribution of risk assessment scores by race/ethnicity. The distribution of scores varies by race/ethnicity most notably for NVCA score, with Black individuals receiving a new violent crime flag proportionately more frequently than White and Hispanic individuals. Table 11 shows the counts associated with the risk distributions from Figure 3. The number of assessed individuals in each race/ethnicity group is sufficient to run statistical tests that look at how the PSA tool scales performed by race/ethnicity.

Figure 3. Distribution of PSA FTA, NCA, and NVCA Risk Scores by Race/Ethnicity

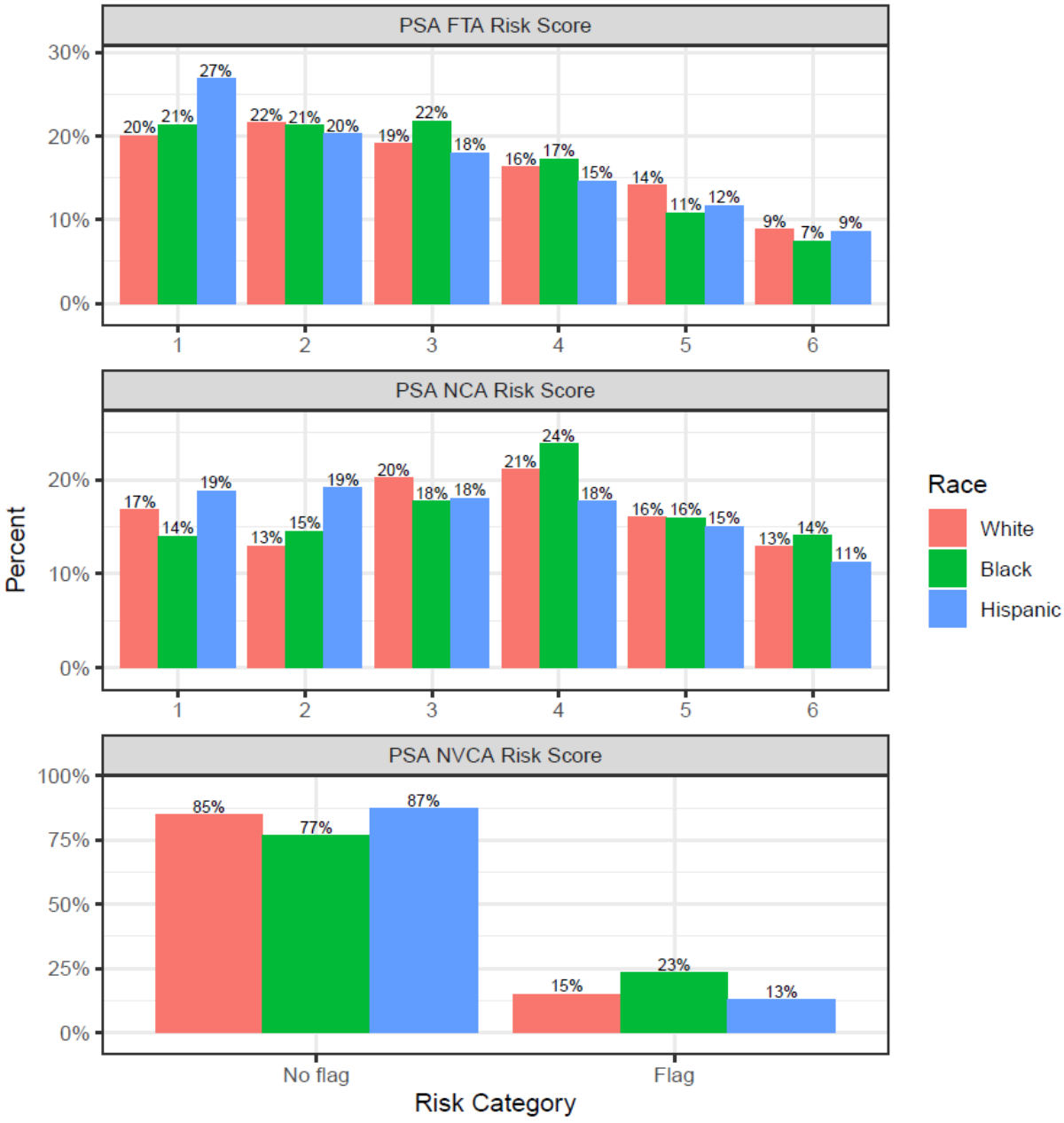


Table 11. Count of Individuals by PSA FTA, NCA, and NVCA Risk Scores and Race/Ethnicity

PSA FTA Risk Score	White	Black	Hispanic
1	1,756	2,044	4,743
2	1,894	2,035	3,582
3	1,673	2,087	3,180
4	1,428	1,651	2,588
5	1,235	1,035	2,074
6	773	710	1,523

PSA NCA Risk Score	White	Black	Hispanic
1	1,474	1,336	3,328
2	1,136	1,390	3,380
3	1,766	1,692	3,190
4	1,849	2,276	3,135
5	1,405	1,525	2,661
6	1,129	1,343	1,996

PSA NVCA Risk Score	White	Black	Hispanic
No flag	7,439	7,336	15,436
Flag	1,320	2,226	2,254

Table 12 shows the AUC values[18] and 95% confidence intervals for each outcome of interest and the relevant PSA risk subscale, separately, for each race/ethnicity group. All AUC values are in the fair to excellent range. Statistical testing[19] indicates that the AUC values for Hispanic individuals are statistically significantly higher compared to both White and Black individuals for the outcomes of FTA, new arrest, new filing, and new conviction. This result indicates that, for these outcomes, the PSA FTA and NCA subscales have a greater ability to distinguish between those who are lower or higher risk for Hispanic individuals than for White or Black individuals.

In addition, the AUC values for Black individuals are statistically significantly lower compared to White individuals for the outcomes of new arrest. This indicates that, for these outcomes, the PSA NCA subscale has a better ability to distinguish between those who are lower or higher risk for White individuals than for Black individuals. All other differences in AUC values are not statistically significant.

[18] See General Validation, for a description of the meaning of AUC values.
[19] See Appendix D for DeLong's test for two ROC curves.

Table 12. AUC Values for Outcomes of Interest by PSA Subscale and Race/Ethnicity

		AUC			CI (95%)		
Risk Score	Outcome	White	Black	Hispanic	White	Black	Hispanic
PSA FTA	FTA	0.667	0.656	0.696	0.655-0.68	0.643-0.669	0.687-0.706
PSA NCA	New Arrest	0.702	0.664	0.739	0.691-0.713	0.654-0.675	0.732-0.746
PSA NCA	New Filing	0.662	0.647	0.702	0.65-0.675	0.633-0.661	0.694-0.711
PSA NCA	New Conviction	0.658	0.641	0.698	0.643-0.673	0.625-0.657	0.689-0.708
PSA NVCA	New Violent Arrest	0.578	0.579	0.575	0.564-0.593	0.565-0.593	0.565-0.585

N White = 8759 , N Black = 9562 , N Hispanic = 17690

Figure 4 shows the results of statistical models of the predictive power of the relevant PSA subscale for each outcome of interest by race/ethnicity group. The lines represent the probability of each outcome of interest at each risk level separately by race/ethnicity. The grey area around each line represents a 95% confidence interval: where the grey areas do not overlap, the evidence indicates there is likely a true difference between the groups; where the grey areas overlap, the evidence may not be strong enough to conclude that there are differences between groups.

For FTA, there is lack of overlap between the confidence intervals over some or all ranges of the tool amongst the Black, Hispanic, and White groups. This indicates that there is evidence that White individuals have a higher probability of FTA than Black individuals with the same score. This is especially apparent at the higher risk levels. There is also evidence that White individuals have a higher probability of FTA than Hispanic individuals with the same score, especially at the midrange of the tool. In addition, at the higher range of the FTA scale, the lack of overlap between the confidence intervals for Black and Hispanic individuals indicates that Hispanic individuals have higher probability of FTA as compared to Black individuals with the same score.

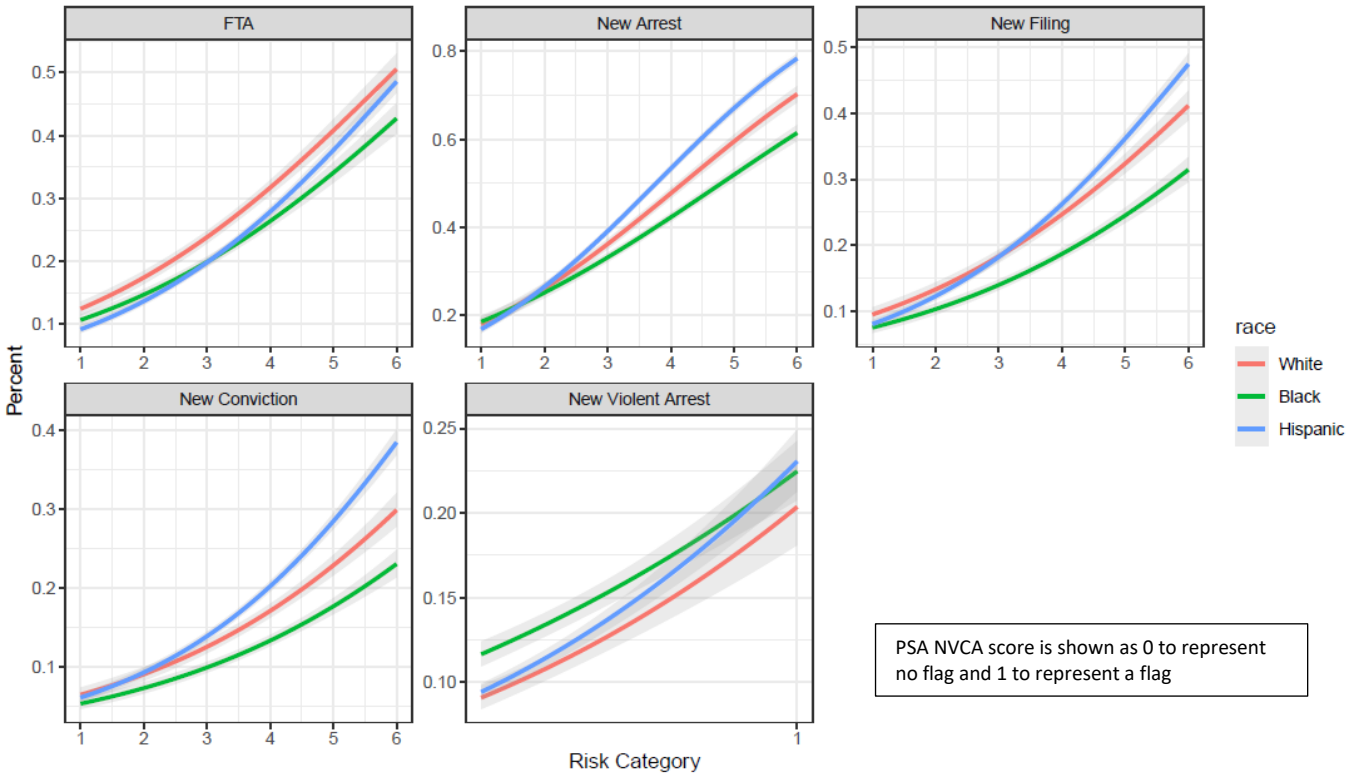
New arrest and new conviction show similar patterns. The confidence intervals for Black, Hispanic, and White individuals do not overlap in the mid-to-high ranges of the scale. This indicates that there is evidence that Hispanic individuals have a higher probability of new arrest and new conviction when compared to Black and White individuals with the same score in the mid-to-high ranges of the scale. Additionally, White individuals scoring in the mid-to-high range of the scale have a higher probability of these outcomes when compared to Black individuals with the same scores.

For new filing, there is lack of overlap between the confidence intervals at the highest range of the scale for Hispanic individuals as compared to White and Black individuals with the same score. This indicates that there is evidence Hispanic individuals have a higher probability of a new filing than Black and White individuals with the same score, especially at the high range of the scale. In addition, White individuals have a higher probability of new filing than Black individuals with the same score across most of the range.

For new violent arrest, the confidence intervals are notably wider since new violent arrests are a relatively rare outcome, which diminishes the ability of the model to make precise predictions. For

individuals without a NVCA flag, there is some evidence that the probability of new violent criminal activity is higher for Black individuals than for White or Hispanic individuals not flagged for NVCA.

Figure 4. Comparison of Racial/Ethnic Differences in Logistic Regression Curves by PSA Outcome of Interest



The patterns of outcomes by risk score and race/ethnicity were also examined across counties using the same tool. One large county showed a similar pattern as above, with all outcomes showing evidence of differences by race/ethnicity across some ranges of the tool, while another large county also showed difference in these outcomes by race/ethnicity, but the differences were smaller and did not significantly vary across the ranges of risk scores.

Table 13 shows the results of a logistic regression that predicts each outcome of interest by the relevant PSA subscale risk score, race, and number of days spent released. Risk level on the relevant PSA subscale is in each case a statistically significant ($p<0.001$) predictor of the outcome of interest. The number of days an individual was out on release ($p<0.001$) also is a statistically significant predictor of all outcomes of interest, indicating that the longer an individual spends on release, the more likely the individual is to experience the outcomes of interest.

This statistical test uses White individuals as the reference group and compares Black and Hispanic individuals to White individuals. Black race is a statistically significant predictor ($p<0.001$) of each outcome of interest. For FTA, new arrest, new filing, and new conviction, Black race has a negative coefficient, indicating that Black individuals had a lower probability of those outcomes as compared to

White individuals with the same risk score. For new violent arrest, Black race has a positive coefficient, indicating that Black individuals had a higher probability of new violent arrest compared to White individuals with the same risk score.

Hispanic ethnicity is a statistically significant predictor of outcome of interest for FTA ($p<0.001$), new arrest ($p<0.001$), new filing ($p<0.05$), and new conviction ($p<0.001$). For FTA, Hispanic ethnicity has a negative coefficient, indicating that Hispanic individuals had a lower probability of FTA as compared to White individuals with the same risk score. For new arrest, new filing, and new conviction, Hispanic ethnicity has a positive coefficient, indicating that Hispanic individuals had a higher probability of these outcomes of interest as compared to White individuals with the same risk score.

This statistical test is limited, however, because it tests for an overall effect of race across the full risk scale. As Figure 4 shows, there are different patterns across particular ranges of the tool subscales. Table 14 uses a more complex statistical model that allows for this possibility.

Table 13. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by PSA Subscale Risk Scores and Race/Ethnicity, Controlling for Days Released

	Dependent variable:				
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest
	(1)	(2)	(3)	(4)	(5)
PSA FTA Risk Score	0.425***				
	(0.008)				
PSA NCA Risk Score		0.551***	0.433***	0.419***	
		(0.008)	(0.009)	(0.010)	
PSA NVCA Risk Score					0.986***
					(0.038)
Race:Black	−0.164***	−0.204***	−0.342***	−0.289***	0.229***
	(0.037)	(0.033)	(0.039)	(0.044)	(0.046)
Race:Hispanic	−0.137***	0.160***	0.070*	0.194***	0.045
	(0.033)	(0.029)	(0.033)	(0.037)	(0.043)
Days Released	0.003***	0.004***	0.003***	0.002***	0.003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	−3.094***	−3.000***	−3.360***	−3.716***	−2.839***
	(0.045)	(0.044)	(0.050)	(0.056)	(0.043)
Observations	36,011	36,011	36,011	36,011	36,011
Log Likelihood	−17,246.450	−20,777.980	−16,877.730	−14,471.760	−12,459.230
Akaike Inf. Crit.	34,502.900	41,565.950	33,765.460	28,953.510	24,928.460

Note: *p<0.05; **p<0.01; ***p<.001

Table 14 shows the results of a logistic regression that predicts each outcome of interest by the relevant PSA subscale risk score, race, the interaction between race and the PSA risk score, and number of days spent released. The number of days spent released is a statistically significant ($p<0.001$) predictor of each outcome of interest. This statistical test again compares Black and Hispanic individuals with White individuals.

The results indicate that there is a statistically significant interaction between Black race and risk score for new arrest indicating that the impact of Black race on new arrest varies at different risk scores. The probability of a new arrest is not statistically significantly different at the lowest end of the scale. However, at the high end of the scale, the probability of new arrest is statistically significantly lower for Black individuals than for White individuals with the same high score.[20]

There is a statistically significant interaction between Hispanic ethnicity and the associated PSA subscale risk score for the outcomes FTA ($p<0.05$), new arrest ($p<0.001$), new filing ($p<0.001$), and new conviction ($p<0.01$), indicating that the effect of Hispanic ethnicity on these outcomes varies at different risk scores. For FTA at the low end of the scale, Hispanic individuals have a statistically significantly lower probability of that outcome relative to White individuals, but at the high end of the scale the probability of FTA converges for Hispanic and White individuals.

For the outcomes of new arrest and new conviction, the difference in the probability of these outcomes for Hispanic individuals scoring at the high end of the scale is statistically significantly higher than it is for White individuals with similarly high scores.

For new filings, the differences in the probability of this outcome are statistically significant at both ends of the scale. At the low end of the scale, Hispanic individuals have a lower probability of a new filing relative to White individuals with the same low score. At the high end of the scale, Hispanic individuals have a higher probability of a new filing relative to White individuals with the same high score.

[20] Throughout this report, regression outputs with interaction terms are shown at the lowest end of the risk scales; main effects were also analyzed at the highest end of the risk scales.

Table 14. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by PSA Subscale Risk Scores, Race/Ethnicity, and Interaction of Race/Ethnicity and Risk Scores, Controlling for Days Released

	Dependent variable:				
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest
	(1)	(2)	(3)	(4)	(5)
PSA FTA Risk Score	0.407***				
	(0.017)				
PSA NCA Risk Score		0.540***	0.402***	0.386***	
		(0.016)	(0.018)	(0.020)	
PSA NVCA Risk Score					1.025***
					(0.080)
Race:Black	−0.109	0.122	−0.199*	−0.141	0.293***
	(0.070)	(0.068)	(0.086)	(0.100)	(0.055)
Race:Hispanic	−0.259***	−0.041	−0.157*	−0.039	0.029
	(0.061)	(0.058)	(0.071)	(0.082)	(0.049)
Days Released	0.003***	0.004***	0.003***	0.002***	0.003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
FTA*Black	−0.023				
	(0.024)				
FTA*Hispanic	0.048*				
	(0.020)				
NCA*Black		−0.116***	−0.044	−0.045	
		(0.022)	(0.026)	(0.029)	
NCA*Hispanic		0.083***	0.078***	0.078**	
		(0.020)	(0.022)	(0.024)	
NVCA*Black					−0.193
					(0.102)
NVCA*Hispanic					0.074
					(0.099)
Constant	−2.620***	−2.421***	−2.833***	−3.195***	−2.849***
	(0.054)	(0.053)	(0.063)	(0.072)	(0.047)
Observations	36,011	36,011	36,011	36,011	36,011
Log Likelihood	−17,239.720	−20,722.660	−16,860.760	−14,458.070	−12,454.270
Akaike Inf. Crit.	34,493.430	41,459.320	33,735.520	28,930.140	24,922.540

Note: *p<0.05; **p<0.01; ***p<.001

Further research is needed to analyze the elements that may be driving the observed differences and whether there are data-driven modifications to the tools' risk factors or weights that can further improve the predictive power of the tool.

GENDER

Figure 5 shows the distribution of risk assessment scores by gender. The charts show a higher percentage of women than men with a low-risk score for FTA and new criminal activity (NCA), and with no flag for new violent criminal activity (NCVA). The number of assessed individuals in each gender group (Table 15) is sufficient to run statistical tests that look at how the PSA tool scales performed by gender.

Figure 5. Distribution of PSA FTA, NCA, and NVCA Risk Scores by Gender

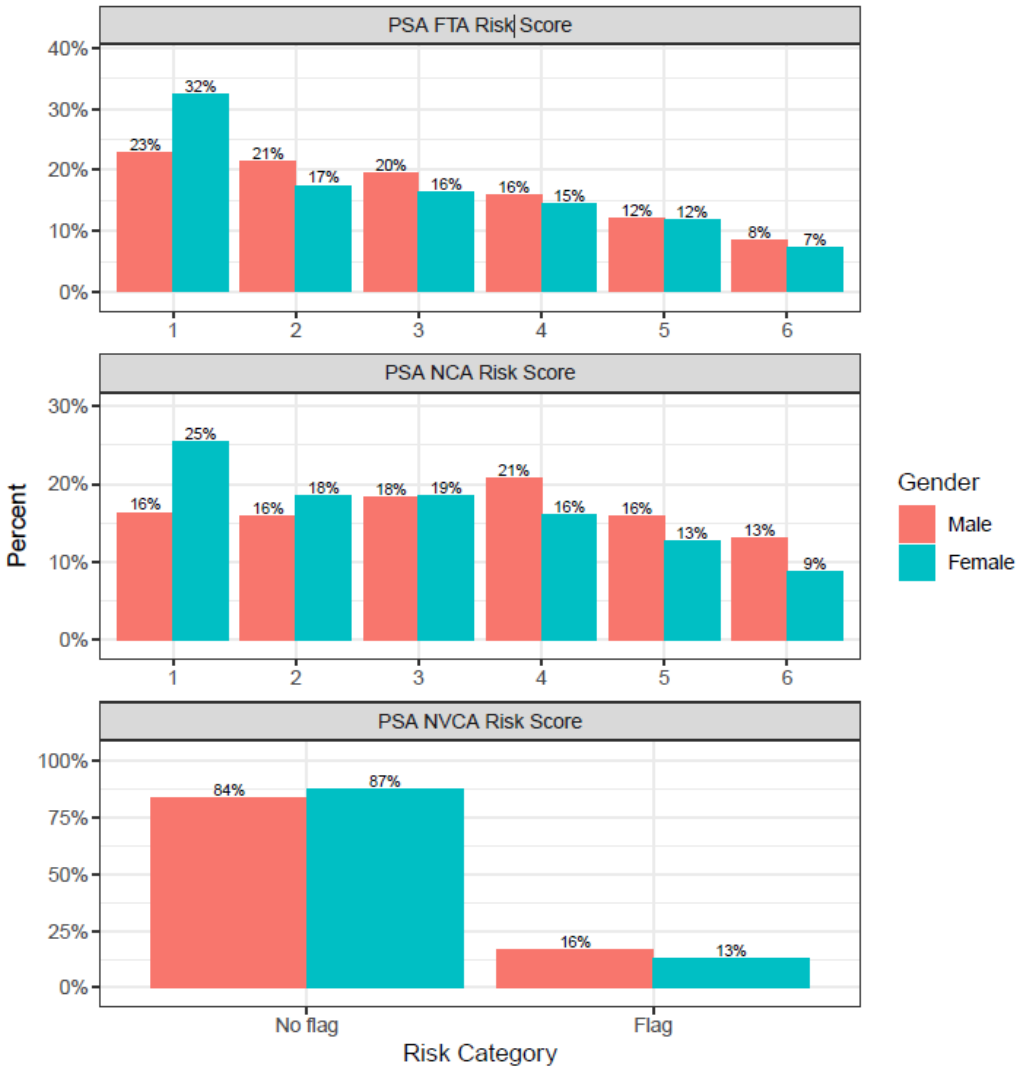


Table 15. Count of Individuals by PSA FTA, NCA, and NVCA Risk Scores and Gender

PSA FTA Risk Score	Male	Female
1	6,982	2,057
2	6,592	1,106
3	6,008	1,036
4	4,871	922
5	3,684	755
6	2,571	461

PSA NCA Risk Score	Male	Female
1	5,020	1,612
2	4,853	1,170
3	5,612	1,178
4	6,358	1,014
5	4,877	805
6	3,988	558

PSA NVCA Risk Score	Male	Female
No flag	25,647	5,536
Flag	5,061	801

Table 16 shows the AUC values[21] and 95% confidence intervals for each outcome of interest, and the relevant PSA risk subscale separately for women and men. AUC values for FTA, new arrest, new filing, and new conviction are in the good to excellent range. AUCs for new violent arrest are in the fair range. Statistical testing[22] indicates that there is no statistically significant difference in AUC between females and males for any of the outcomes of interest.

Table 16. AUC Values for Outcomes of Interest by PSA Subscale and Gender

		AUC		CI (95%)	
Risk Score	Outcome	Female	Male	Female	Male
PSA FTA	FTA	0.693	0.684	0.678-0.709	0.677-0.691
PSA NCA	New Arrest	0.710	0.711	0.697-0.723	0.705-0.716
PSA NCA	New Filing	0.674	0.678	0.657-0.691	0.672-0.685
PSA NCA	New Conviction	0.675	0.672	0.656-0.694	0.665-0.68
PSA NVCA	New Violent Arrest	0.582	0.579	0.565-0.6	0.572-0.587

N Female = 6337 , N Male = 30708

[21] See General Validation for a description of the meaning of AUC values.
[22] See Appendix D for DeLong's test for two ROC curves.

Figure 6 shows the results of statistical models comparing the predictive power of the relevant PSA subscale for each outcome of interest for women and men. The lines represent the probability of each outcome of interest at each risk level separately for each gender. The grey area around each line represents a 95% confidence interval: where the grey areas do not overlap, the evidence indicates that there is likely a true difference between the groups; where the grey areas overlap, the evidence may not be strong enough to conclude that there are differences between them.

For the outcomes of new filings and new convictions, there is a lack of overlap between the female and male confidence intervals in the mid-to-high ranges of the tool.[23] This indicates that there is evidence that the probability of these outcomes is lower for women relative to men with the same PSA NCA risk score at those ranges. For the outcome of FTA, the confidence intervals overlap, indicating that there may not be a statistically significant difference between females and males for this outcome. For new violent arrest, women show a lower probability of this outcome when the PSA NVCA subscale indicates no new violent flag. When the tool indicates a violent flag, however, women show a higher rate of new violent arrest, though the wide confidence interval indicates that, given the rarity of a violent flag among women combined with the low rates of new violent arrest overall, there may not be enough evidence to reliably indicate a true difference.

Figure 6. Comparison of Gender Differences in Logistic Regression Curves by PSA Outcome of Interest

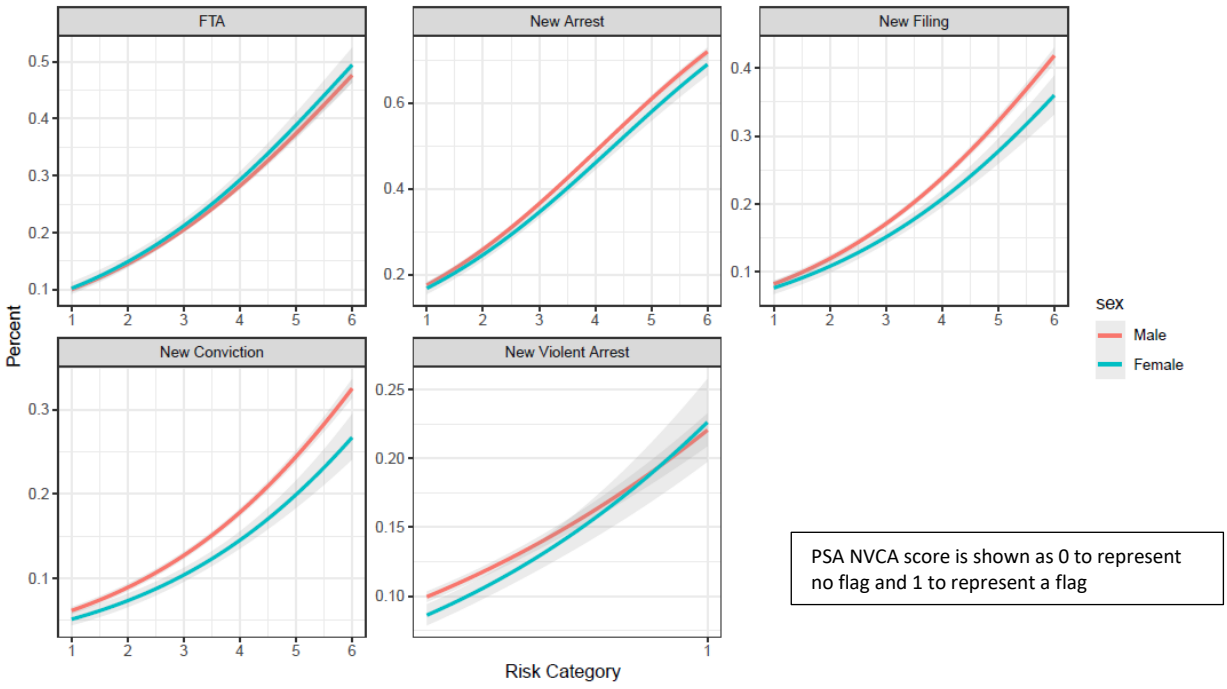


Table 17 shows the results of a logistic regression that predicts each outcome of interest by the relevant PSA subscale risk score, number of days spent released, and gender. This statistical test compares

[23] The confidence intervals for new arrests appear to be overlapping in Figure 6, but the logistic regression detects a significant difference with women having a lower probability of new arrest relative to men with the same risk score.

women with the base group of men. The number of days the individual was out on release is a statistically significant predictor (p<0.001) of all the outcomes of interest, indicating that the longer an individual spends on release the more likely the individual is to experience the outcomes of interest.

Female gender is a statistically significant predictor of new arrest (p<0.001), new filing (p<0.001), new conviction (p<0.001), and new violent arrest (p<0.05), indicating that for each outcome, women are statistically significantly less likely to experience these outcomes relative to men with the same risk score.

This statistical test is limited, however, because it tests for an overall effect of gender across the full risk scale and, as can be seen from Figure 6, there are to be different patterns for women than for men, especially for the new violent arrest outcome. Table 18 uses a more complex statistical model that allows for this possibility.

Table 17. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by PSA Subscale Risk Scores and Gender, Controlling for Days Released

	Dependent variable:				
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest
	(1)	(2)	(3)	(4)	(5)
PSA FTA Risk Score	0.439***				
	(0.008)				
PSA NCA Risk Score		0.548***	0.427***	0.410***	
		(0.008)	(0.009)	(0.010)	
PSA NVCA Risk Score					1.030***
					(0.038)
Female	0.035	−0.123***	−0.169***	−0.256***	−0.115*
	(0.036)	(0.032)	(0.038)	(0.043)	(0.045)
Days Released	0.003***	0.004***	0.003***	0.003***	0.003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	−3.274***	−2.985***	−3.391***	−3.649***	−2.794***
	(0.040)	(0.038)	(0.044)	(0.049)	(0.031)
Observations	37,045	37,045	37,045	37,045	37,045
Log Likelihood	−17,545.210	−21,273.730	−17,263.530	−14,778.650	−12,754.730
Akaike Inf. Crit.	35,098.420	42,555.460	34,535.060	29,565.290	25,517.460
Note:					*p<0.05; **p<0.01; ***p<.001

Table 18 shows the results of a logistic regression that predicts each outcome of interest by the relevant PSA subscale risk score, gender, the interaction between gender and the PSA risk score, and number of

days spent released. The number of days spent released is a statistically significant predictor (p<0.001) of each outcome of interest. This statistical test again compares women with men as the base group.

The results indicate that there is a statistically significant interaction between gender and risk score on new violent arrest, indicating that the impact of gender on new violent offense varies at different risk scores. The results indicate that women have statistically significantly lower probability of new filing at a PSA NCA score of one, relative to men with the same score. However, at the high end of the scale, differences in the probability of a new violent arrest are not statistically significant between men and women.

Table 18. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by PSA Subscale Risk Scores, Gender, and Interaction of Gender and Risk Scores, Controlling for Days Released

	Dependent variable:				
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest
	(1)	(2)	(3)	(4)	(5)
PSA FTA Risk Score	0.437***				
	(0.009)				
PSA NCA Risk Score		0.550***	0.431***	0.412***	
		(0.009)	(0.010)	(0.011)	
PSA NVCA Risk Score					0.995***
					(0.041)
Female	0.006	−0.107	−0.098	−0.211*	−0.173***
	(0.063)	(0.057)	(0.074)	(0.087)	(0.052)
Days Released	0.003***	0.004***	0.003***	0.003***	0.003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
FTA*Female	0.012				
	(0.022)				
NCA*Female		−0.007	−0.026	−0.016	
		(0.020)	(0.024)	(0.027)	
NVCA*Female					0.246*
					(0.106)
Constant	−2.829***	−2.439***	−2.976***	−3.247***	−2.786***
	(0.035)	(0.033)	(0.038)	(0.043)	(0.031)
Observations	37,045	37,045	37,045	37,045	37,045
Log Likelihood	−17,545.050	−21,273.670	−17,262.920	−14,778.480	−12,752.070
Akaike Inf. Crit.	35,100.090	42,557.330	34,535.830	29,566.950	25,514.140

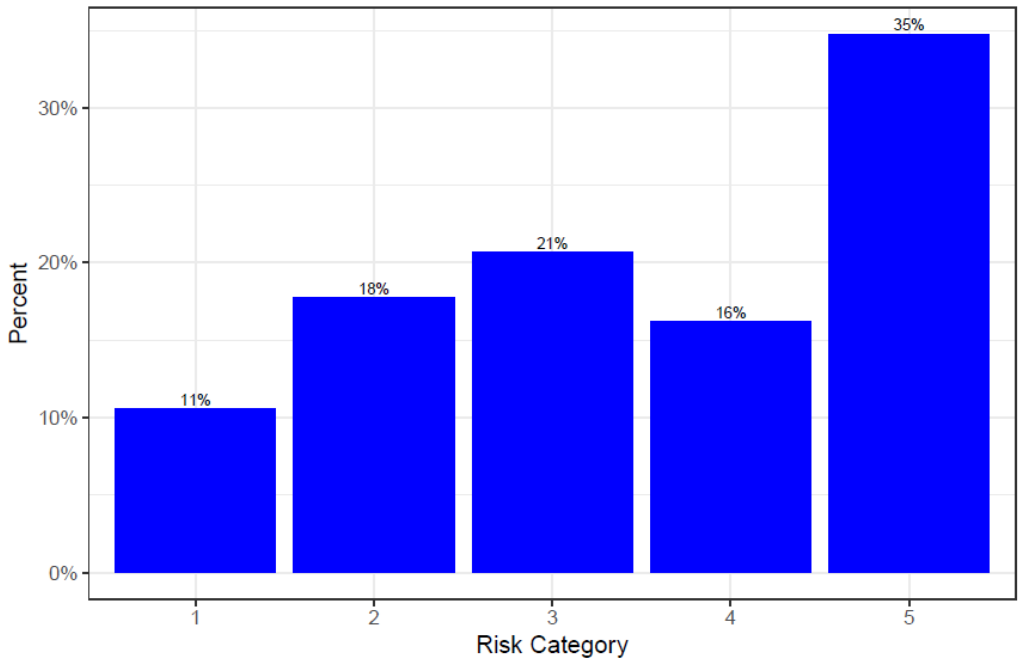
Note: *p<0.05; **p<0.01; ***p<.001

VIRGINIA PRETRIAL RISK ASSESSMENT INSTRUMENT (VPRAI) VALIDATION

GENERAL VALIDATION

Figure 7 shows the distribution of risk categories for individuals in the evaluation data set assessed with the VPRAI tool used by the Pretrial Pilot Project in San Joaquin County. The VPRAI tool developer divided the risk scores into five risk levels: risk level 1 includes scores 0–1, level 2 includes score 2, level 3 includes score 3, level 4 includes score 4, and level 5 includes scores 5–9.[9] Risk level 1 was the least common assessed risk level in the evaluation data set. Risk level 4 was less common than levels 2 and 3, and level 5 was the most assessed risk level. The VPRAI was specifically designed to predict a composite of failure to appear (FTA), new arrests (NA), and technical violations.[24] The distribution of all assessed individuals may differ from the distribution in the evaluation data set because the evaluation data set includes only released individuals with concluded pretrial periods. Table 19 shows the counts associated with the risk distribution in Figure 7.

Figure 7. Distribution of VPRAI Risk Categories



[24] The low quality of the data on technical violations prevented us from creating a composite failure rate that included FTA, new arrest, and technical violations. The composite measure in this report is a combination of the risk of FTA and the risk of new arrest.

Table 19. Counts of Individuals by VPRAI Risk Levels

VPRAI Risk Level	Total
1	600
2	1,008
3	1,176
4	922
5	1,975

Table 20 shows the AUC values for the VPRAI tool, using the five established risk levels for each outcome of interest. The AUC value is a single number representing the tool's ability to discriminate between individuals at lower or higher risk across the range of the tool. For criminal justice risk assessments, a common metric for evaluating AUC values is derived from Desmarais and Singh (2013),[25] who defined AUC values less than 0.55 as poor, 0.55–0.63 as fair, 0.64–0.70 as good, and 0.71–1.00 as excellent. By these definitions, the AUC values for all outcomes fall into the good to excellent range.

The 95% confidence interval—which represents the range of AUC estimates that the true AUC value is 95% likely to fall between—is also shown. A smaller range indicates that given the sample size and pattern of the data, the AUC can be estimated with greater precision. None of the 95% confidence intervals fall below the fair range.

Table 20. AUC Values by Outcomes of Interest and VPRAI Risk Levels

Outcome	AUC	CI (95%)
FTA	0.692	0.678-0.706
New Arrest	0.745	0.732-0.757
New Filing	0.705	0.691-0.718
New Conviction	0.695	0.68-0.709
New Violent Arrest	0.643	0.627-0.66
FTA or New Arrest	0.741	0.729-0.754

N = 5681

Figure 8 shows the rate of various adverse outcomes during the pretrial period at each risk level of the VPRAI.[26] For each outcome of interest,[27] observed rates of the outcome increase as the assessed risk level increases.

Figure 8. VPRAI Outcomes by Risk Category

[25] Sarah L. Desmarais and Jay P. Singh, *Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States* (Lexington, KY: Council of State Governments, 2013).
[26] Risk levels are groupings of scores as defined by the tool developer.
[27] See the validation methodology section for definitions of each outcome of interest.

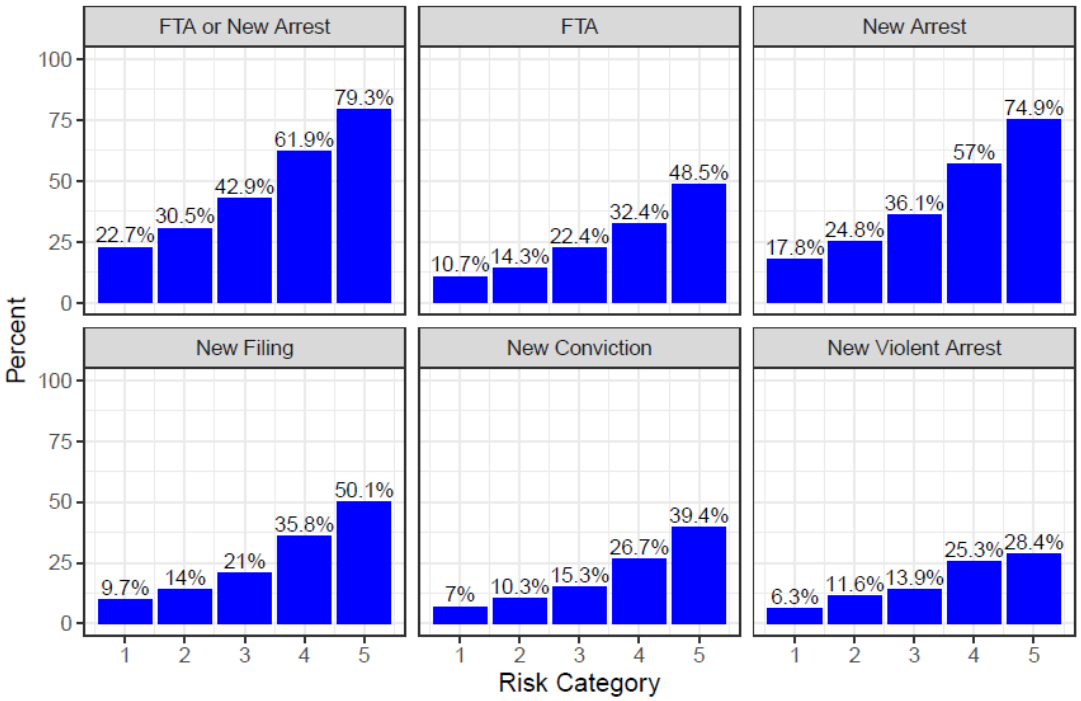


Table 21 shows the results from the logistic regression model predicting each outcome of interest, the model control for the number of days the defendant spent released during the pretrial period. For each outcome of interest, the model shows that the association between the VPRAI risk score and the likelihood of the outcome during the pretrial period is statistically significant ($p<0.001$), as is the number of days released.

Table 21. Logistic Regression Models Predicting the Likelihood of Outcomes of Interest by VPRAI Risk Scores Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
VPRAI Risk Score	0.571***	0.758***	0.634***	0.597***	0.410***	0.738***
	(0.026)	(0.025)	(0.027)	(0.029)	(0.028)	(0.024)
Days Released	0.002***	0.003***	0.003***	0.002***	0.002***	0.003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	−3.681***	−3.526***	−4.022***	−4.239***	−3.731***	−3.236***
	(0.116)	(0.108)	(0.122)	(0.133)	(0.128)	(0.104)
Observations	5,681	5,681	5,681	5,681	5,681	5,681
Log Likelihood	−2,993.521	−3,108.618	−2,910.796	−2,642.144	−2,504.607	−3,090.460
Akaike Inf. Crit.	5,993.043	6,223.237	5,827.591	5,290.288	5,015.214	6,186.920
Note:						*p<0.05; **p<0.01; ***p<.001

ANALYSIS OF PREDICTIVE BIAS

RACE/ETHNICITY

Figure 9 shows the distribution of risk assessment scores by race/ethnicity. Although the distribution of scores for the individual race/ethnicity groups is consistent with the overall distribution (Figure 7), with risk level 1 the least common and risk level 5 the most common assessed risk level, the White race group had the largest shares of individuals in both the highest and lowest risk levels. Table 22 shows the counts associated with the risk distributions from Figure 9. The number of assessed individuals in each race/ethnic group (Table 22) is sufficient to run statistical tests that look at how the tool performed by race/ethnicity.

Figure 9. Distribution of VPRAI Risk Scores by Race/Ethnicity

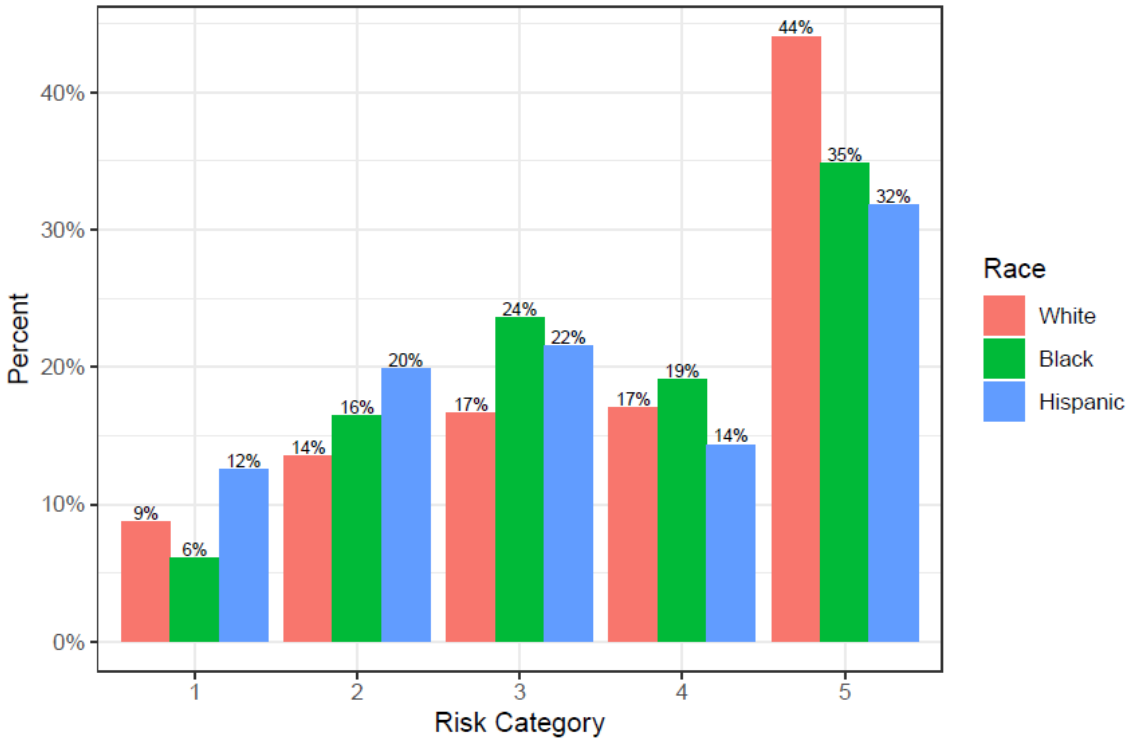


Table 22. Count of Individuals by VPRAI Risk Scores and Race/Ethnicity

VPRAI Risk Score	White	Black	Hispanic
1	138	77	286
2	214	208	454
3	263	299	492
4	269	242	328
5	697	441	728

Table 23 shows the AUC values[28] and 95% confidence intervals for each outcome of interest separately for each race/ethnic group. With the exception of the new violent arrest AUC values for White and Black individuals, which fall in the fair range, most AUC values are in the good to excellent range. The AUC values for the Black individuals are lower relative to the AUC values for White and Hispanic individuals across all outcomes. Statistical testing[29] indicates that the AUC values of Black individuals are statistically significantly lower than the AUC values of White individuals for FTA, new arrest, and the combined measure of "FTA or New Arrest." The AUC values of Black individuals are also statistically significantly lower than the AUC values of Hispanic individuals across all outcomes of interest. This result indicates

[28] See General Validation for a description of the meaning of AUC values.
[29] See Appendix D for DeLong's test for two ROC curves.

that, for Black individuals, the VPRAI has a poorer ability to distinguish between individuals who are lower or higher risk for some outcomes. In addition, findings show that the AUC values of White individuals are statistically significantly lower than the AUC values of Hispanic individuals for new filing, new conviction, and new violent arrest. This indicates that relative to Hispanic individuals, for White individuals the VPRAI has a poorer ability to distinguish between those at lower or higher risk for some outcomes.

Table 23. AUC Values for the VPRAI, by Outcomes of Interest and Race/Ethnicity

	AUC			CI (95%)		
Outcome	White	Black	Hispanic	White	Black	Hispanic
FTA	0.697	0.647	0.689	0.672-0.721	0.616-0.678	0.666-0.712
New Arrest	0.748	0.703	0.749	0.725-0.772	0.675-0.73	0.729-0.768
New Filing	0.681	0.670	0.722	0.656-0.706	0.64-0.7	0.701-0.744
New Conviction	0.670	0.664	0.713	0.643-0.696	0.631-0.696	0.69-0.736
New Violent Arrest	0.608	0.628	0.669	0.576-0.64	0.595-0.66	0.644-0.695
FTA or New Arrest	0.760	0.697	0.737	0.736-0.783	0.669-0.725	0.717-0.757

N White = 1581 , N Black = 1267 , N Hispanic = 2288

Figure 10 shows the results of statistical models of the predictive power of the tool for each race/ethnic group. Each line represents the probability of each outcome of interest at each risk level separately for each race/ethnicity. The grey area around each line represents a 95% confidence interval: where the grey areas do not overlap, the evidence indicates that there is likely a true difference between the groups; where the grey areas overlap, the evidence is not strong enough to conclude that there are differences between them.

For the combined measure of "FTA or New Arrest," FTA, new arrest, new filing, and new conviction, the confidence intervals overlap, indicating that there may not be evidence of a difference in the likelihood of those outcomes for individuals of different race/ethnic groups with the same score.

Due to the rarity of this outcome, the confidence intervals are notably wider for new violent arrest, diminishing the ability of the model to make precise predictions. Nevertheless, Figure 10 shows that the confidence intervals of Black individuals do not overlap with the confidence intervals of Hispanic and White individuals at the midrange of the tool. This indicates that, at the midrange of the tool, Black individuals have a higher probability of new violent arrest than White or Hispanic individuals with similar scores. Furthermore, at the high range of the tool, the confidence interval for White individuals does not overlap with the confidence intervals of Black and Hispanic individuals indicating that White individuals have a lower probability of a new violent arrest than Black and Hispanic individuals with the same scores.

Figure 10. Comparison of Racial/Ethnic Differences in Logistic Regression Curves by VPRAI Outcome of Interest

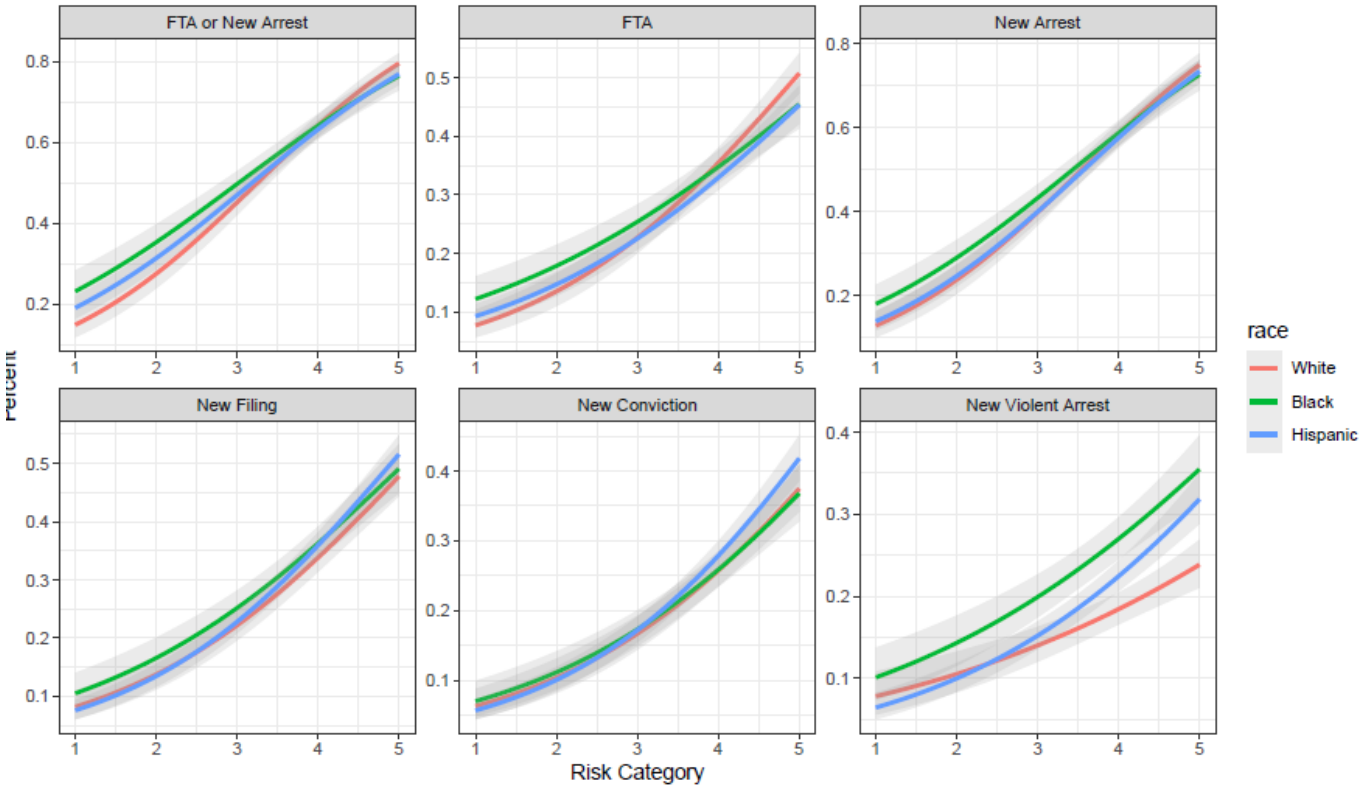


Table 24 shows the results of a logistic regression that predicts each outcome of interest by VPRAI risk score, the number of days spent released, and race. The risk score is a statistically significant predictor of each outcome of interest ($p<0.001$). The number of days the individual was out on release also is a statistically significant predictor of all the outcomes of interest, indicating that the longer an individual spends on release, the more likely the individual is to experience the outcomes of interest.

This statistical test compares Black and Hispanic individuals using White individuals as a reference group. Black and Hispanic race/ethnicity were statistically significant ($p<0.001$) predictors of new violent arrest, with positive coefficients indicating a greater likelihood of new violent arrest for Black and Hispanic individuals compared to White individuals with the same risk score. Race/ethnicity was not a statistically significant predictor of any other outcomes of interest. This statistical test is limited, however, because it tests for an overall effect of race across the full risk scale. As can be seen from Figure 10, some differences may emerge only in particular ranges of the tool. Table 25 uses a more complex statistical model that allows for this possibility.

Table 24. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by VPRAI Risk Scores and Race/Ethnicity, Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
VPRAI Risk Score	0.550***	0.751***	0.630***	0.593***	0.416***	0.730***
	(0.027)	(0.026)	(0.028)	(0.031)	(0.030)	(0.026)
Race:Black	−0.072	0.020	0.086	−0.032	0.513***	0.045
	(0.087)	(0.088)	(0.090)	(0.096)	(0.098)	(0.088)
Race:Hispanic	−0.138	−0.040	0.090	0.115	0.266**	−0.007
	(0.077)	(0.077)	(0.079)	(0.084)	(0.090)	(0.077)
Days Released	0.002***	0.003***	0.003***	0.002***	0.002***	0.003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	−3.513***	−3.486***	−4.079***	−4.274***	−3.962***	−3.209***
	(0.135)	(0.129)	(0.144)	(0.156)	(0.155)	(0.126)
Observations	5,136	5,136	5,136	5,136	5,136	5,136
Log Likelihood	−2,745.887	−2,813.810	−2,658.855	−2,422.353	−2,311.731	−2,790.987
Akaike Inf. Crit.	5,501.774	5,637.621	5,327.711	4,854.707	4,633.462	5,591.974
Note:						*p<0.05; **p<0.01; ***p<.001

Table 25 shows the results of a logistic regression that predicts each outcome of interest by VPRAI risk score, race, the interaction between race and the VPRAI risk score, and number of days spent released. This statistical test compares Black and Hispanic individuals with White individuals as a reference group.

There is a statistically significant interaction between Black race and risk score for FTA ($p<0.01$), new arrest ($p<0.05$), and "FTA or New Arrest" ($p<0.01$). The statistically significant interaction indicates that the impact of Black race on the outcomes of interest varies at different risk scores. At the low end of the FTA scale, Black individuals have a statistically significantly higher likelihood of FTA than White individuals with the same score, while at the high end of the scale Black individuals have a statistically significantly lower likelihood of FTA when compared with White individuals with the same score. For new arrest, the pattern is similar, but the difference is only statistically significant at the low end of the scale. For the composite outcome of "FTA or New Arrest," Black race is predictive of a statistically significantly greater likelihood of "FTA or New Arrest" at the low end of the scale and a statistically significantly greater likelihood of "FTA or New Arrest" at the high end of the scale when compared with White individuals with the same scores.

There is also a statistically significant interaction between Hispanic ethnicity and risk score for FTA and "FTA or New Arrest." The significant interaction indicates that the impact of Hispanic ethnicity on these outcomes varies at different risk scores. For FTA, Hispanic individuals at the high end of the scale are less likely to have an outcome of FTA than White individuals with the same scores, but the difference is not statistically significant. In contrast, Hispanic individuals at the low end of the scale have a statistically significantly higher probability of "FTA or New Arrest" relative to White individuals with the same score.

Table 25. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by VPRAI Risk Scores, Race/Ethnicity, and Interaction of Race/Ethnicity and Risk Scores, Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
VPRAI Risk Score	0.666***	0.820***	0.636***	0.586***	0.336***	0.847***
	(0.052)	(0.048)	(0.053)	(0.057)	(0.057)	(0.048)
Race:Black	0.581*	0.473*	0.347	0.168	0.335	0.640**
	(0.243)	(0.214)	(0.250)	(0.280)	(0.265)	(0.207)
Race:Hispanic	0.283	0.142	0.002	−0.020	−0.144	0.361*
	(0.211)	(0.182)	(0.216)	(0.238)	(0.236)	(0.175)
Days Released	0.002***	0.003***	0.003***	0.002***	0.002***	0.003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
VPRAI*Black	−0.217**	−0.166*	−0.089	−0.066	0.055	−0.226**
	(0.076)	(0.072)	(0.077)	(0.085)	(0.082)	(0.071)
VPRAI*Hispanic	−0.137*	−0.064	0.032	0.046	0.136	−0.138*
	(0.065)	(0.061)	(0.066)	(0.072)	(0.072)	(0.060)
Constant	−3.329***	−2.931***	−3.466***	−3.656***	−3.297***	−2.802***
	(0.182)	(0.159)	(0.186)	(0.203)	(0.199)	(0.155)
Observations	5,136	5,136	5,136	5,136	5,136	5,136
Log Likelihood	−2,741.488	−2,811.101	−2,657.366	−2,421.291	−2,309.877	−2,785.666
Akaike Inf. Crit.	5,496.977	5,636.202	5,328.732	4,856.583	4,633.754	5,585.332
Note:						*p<0.05; **p<0.01; ***p<.001

GENDER

Figure 11 shows the distribution of risk assessment scores by gender. The distribution of scores for the individual gender groups is consistent with the overall distribution (Figure 7), with risk level 1 the least common and risk level 5 the most common assessed risk levels. However, the distribution for males relative to females was skewed toward the higher risk levels. Table 26 shows the counts associated with the risk distributions from Figure 11, and indicates that the number of individuals in each gender group is sufficient to run statistical tests that look at how the VPRAI tool scales performed by gender.

Figure 11. Distribution of VPRAI Risk Level by Gender

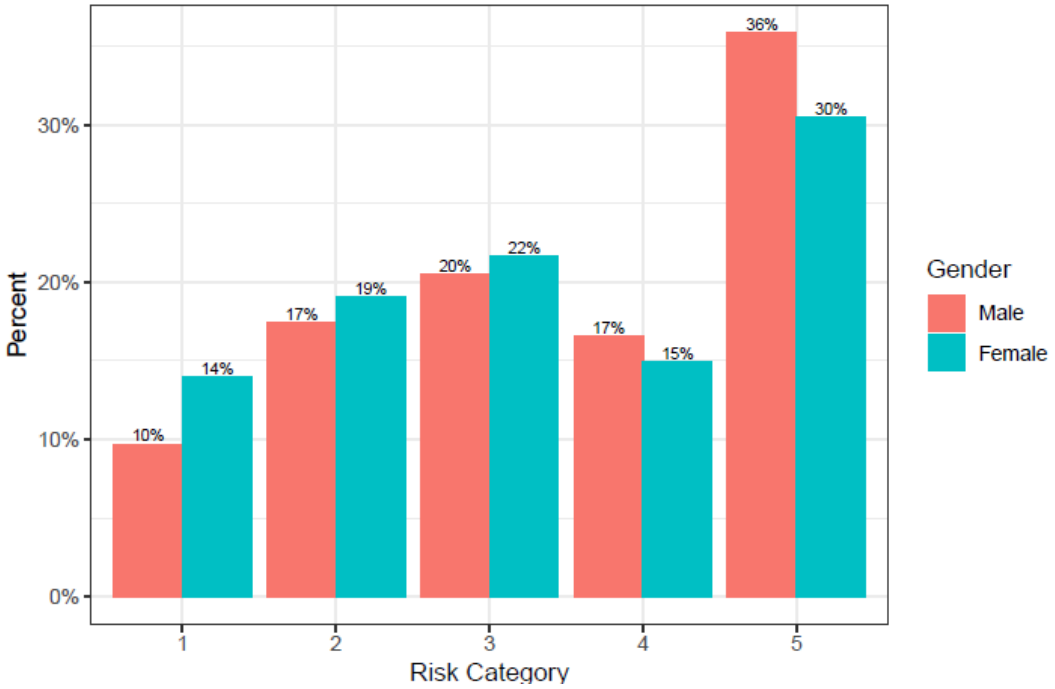


Table 26. Count of Individuals by VPRAI Risk Score and Gender

VPRAI Risk Score	Male	Female
1	442	158
2	792	216
3	931	245
4	753	169
5	1,630	345

Table 27 shows the AUC values[16] and 95% confidence intervals for each outcome of interest and VPRAI risk score, separately for women and men. Except for the new violent arrest AUC value for women, which falls in the fair range, all other AUC values are in the good to excellent range. Statistical testing[30] indicates that there is no evidence of a statistically significant difference between the AUCs for women as compared to men.

[30] See Appendix D for DeLong's test for two ROC curves.

Table 27. AUC Values for the VPRAI, by Outcome of Interest and Gender

	AUC		CI (95%)	
Outcome	Female	Male	Female	Male
FTA	0.687	0.694	0.655-0.719	0.678-0.709
New Arrest	0.721	0.749	0.691-0.75	0.736-0.763
New Filing	0.683	0.708	0.649-0.717	0.693-0.723
New Conviction	0.665	0.698	0.626-0.705	0.682-0.714
New Violent Arrest	0.624	0.644	0.579-0.669	0.626-0.661
FTA or New Arrest	0.725	0.744	0.696-0.754	0.73-0.758

N Female = 1133 , N Male = 4548

Figure 12 shows the results of statistical models of the predictive power of the relevant VPRAI subscale for each outcome of interest for women and men. Each line represents the probability of each outcome of interest at each risk level separately by gender. The grey area around each line represents a 95% confidence interval: where the grey areas do not overlap, the evidence indicates that a true difference between the groups is likely; where the grey areas overlap, the evidence may not be strong enough to conclude that there are differences between them.

For all outcomes except FTA, there is a lack of overlap between confidence intervals for men and women in the mid-to-high ranges of the tool. This indicates that there is evidence that the probability of these outcomes is lower for women relative to men with the same risk score at those ranges. For the outcome of FTA, the confidence intervals overlap, indicating that there may not be a statistically significant difference between women and men for this outcome.

Figure 12. Comparison of Gender Differences in Logistic Regression Curves by VPRAI Outcome of Interest

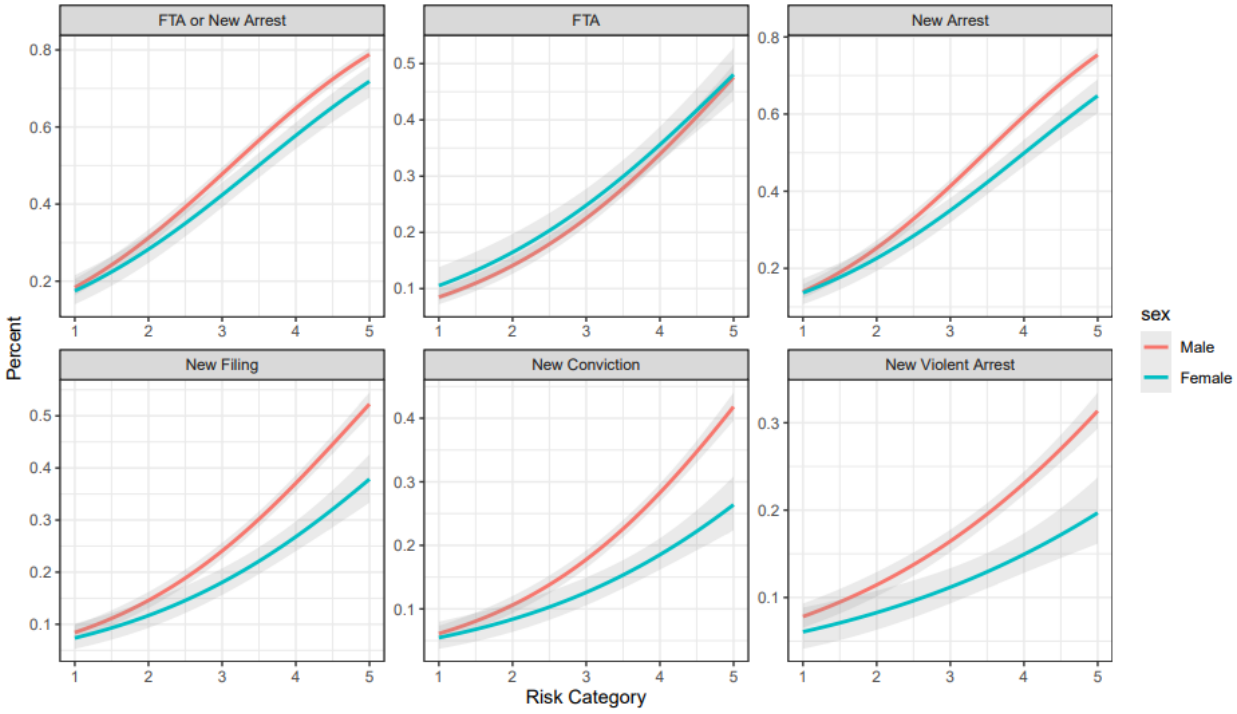


Table 28 shows the results of a logistic regression that predicts each outcome of interest by the relevant VPRAI risk score, gender, and number of days spent released. This statistical test compares women with the base group of men. Risk level on the relevant VPRAI score is a statistically significant predictor of all outcomes of interest ($p<0.001$). The number of days the individual was out on release also is a statistically significant predictor of all the outcomes of interest, indicating that the longer an individual spends on release, the more likely the individual is to experience the outcomes of interest.

Female gender is a statistically significant predictor ($p<0.001$) of all the outcomes of interest except for FTA, indicating that for new arrest, new filing, new conviction, new violent arrest, and "FTA or New Arrest," women are statistically significantly less likely to experience the outcome of interest than men with the same risk score.

This statistical test is limited, however, because it tests for an overall effect of gender across the full risk scale, and, as can be seen from Figure 12, patterns may be different for women than men across different ranges of the tool. Table 29 uses a more complex statistical model that allows for this possibility.

Table 28. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by VPRAI Risk Scores and Gender, Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
VPRAI Risk Score	0.574***	0.755***	0.629***	0.590***	0.402***	0.736***
	(0.026)	(0.025)	(0.027)	(0.029)	(0.029)	(0.024)
Female	0.111	−0.357***	−0.490***	−0.572***	−0.546***	−0.272***
	(0.080)	(0.078)	(0.086)	(0.096)	(0.101)	(0.077)
Days Released	0.002***	0.003***	0.003***	0.002***	0.002***	0.003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	−3.713***	−3.448***	−3.917***	−4.119***	−3.612***	−3.174***
	(0.119)	(0.109)	(0.123)	(0.134)	(0.129)	(0.106)
Observations	5,681	5,681	5,681	5,681	5,681	5,681
Log Likelihood	−2,992.551	−3,098.118	−2,894.020	−2,623.057	−2,488.781	−3,084.275
Akaike Inf. Crit.	5,993.102	6,204.235	5,796.040	5,254.114	4,985.561	6,176.550
Note:						*p<0.05; **p<0.01; ***p<.001

Table 29 shows the results of a logistic regression that predicts each outcome of interest by VPRAI risk score, gender, the interaction between gender and the VPRAI risk score, and number of days spent released. The number of days spent released is a statistically significant predictor of each outcome of interest. This statistical test again compares women with men as the base group.

The results show no statistically significant interaction between gender and risk score for any outcome. In the absence of a statistically significant interaction, the model with no interaction shown in Table 28 is more appropriate to demonstrate the impact of gender.

Table 29. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by VPRAI Risk Scores, Gender, and Interaction of Gender and Risk Scores, Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
VPRAI Risk Score	0.578***	0.772***	0.644***	0.611***	0.413***	0.743***
	(0.029)	(0.028)	(0.030)	(0.032)	(0.031)	(0.027)
Female	0.167	−0.141	−0.260	−0.203	−0.355	−0.183
	(0.187)	(0.170)	(0.210)	(0.237)	(0.235)	(0.161)
Days Released	0.002***	0.003***	0.003***	0.002***	0.002***	0.003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
VPRAI*Female	−0.020	−0.084	−0.081	−0.126	−0.068	−0.037
	(0.062)	(0.059)	(0.068)	(0.075)	(0.077)	(0.058)
Constant	−3.152***	−2.735***	−3.329***	−3.592***	−3.242***	−2.456***
	(0.104)	(0.093)	(0.106)	(0.116)	(0.112)	(0.090)
Observations	5,681	5,681	5,681	5,681	5,681	5,681
Log Likelihood	−2,992.498	−3,097.118	−2,893.315	−2,621.689	−2,488.391	−3,084.076
Akaike Inf. Crit.	5,994.996	6,204.236	5,796.629	5,253.378	4,986.783	6,178.151

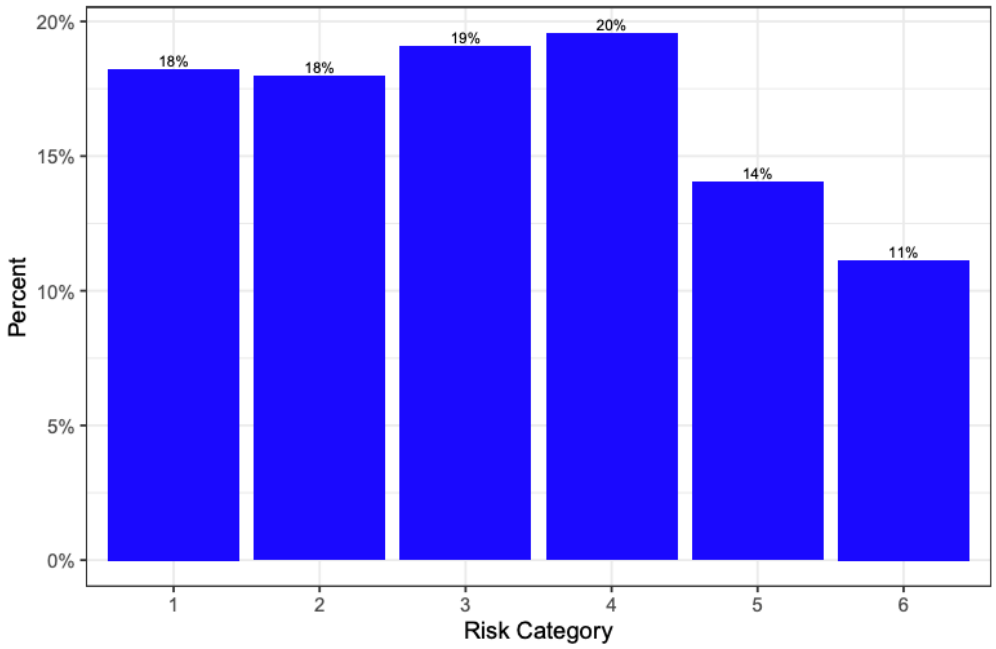
Note: *p<0.05; **p<0.01; ***p<.001

VIRGINIA PRETRIAL RISK ASSESSMENT INSTRUMENT, REVISED (VPRAI-R) VALIDATION

GENERAL VALIDATION

Figure 13 shows the distribution of risk levels for individuals in the evaluation data set assessed with the VPRAI-R tool, used by the Pretrial Pilot Projects in Alameda and San Mateo Counties. The VPRAI-R tool developer divided the risk scores into six risk categories: level 1 includes scores 0–2; level 2, scores 3–4; level 3, scores 5–6; level 4, scores 7–8; level 5, scores 9–10; and level 6, scores 11–14. Most assessments fell into risk levels 1–4, and level 6 was the least common in the evaluation data set. The VPRAI-R was specifically designed to predict a composite of failure to appear (FTA), new arrest (NA), and technical violations.[31] The distribution of all assessed individuals may differ from the distribution in the evaluation data set because the evaluation data set includes only released individuals with concluded pretrial periods. Table 30 shows the counts associated with the risk distributions from Figure 13.

Figure 13. Distribution of VPRAI-R Risk Scores



[31]The low quality of the data on technical violations prevented the creation of a composite failure rate that included FTA, new arrest, and technical violations. The composite measure in this report is a combination of the risk of FTA and the risk of new arrest.

Table 30. Counts of Individuals by VPRAI-R Risk Scores

VPRAIR Risk Level	Total
1	727
2	717
3	761
4	780
5	560
6	444

Table 31 shows the AUC values for the VPRAI-R tool, using the six established risk levels, for each outcome of interest. The AUC value is a single number that represents the ability of the tool to discriminate between individuals who are lower or higher risk across the range of the tool. For criminal justice risk assessments, a common metric for evaluating AUC values is derived from Desmarais and Singh (2013),[32] who defined AUC values less than 0.55 as poor, 0.55–0.63 as fair, 0.64–0.70 as good, and 0.71–1.00 as excellent. By these definitions, the AUC values for the VPRAI-R are excellent for new arrest and the composite measure of "FTA or New Arrest;" good for FTA, new filing, and new conviction; and fair for new violent arrest.

The 95% confidence interval—which represents the range of AUC estimates that the true AUC value is 95% likely to fall between—is also shown. A smaller range indicates that, given the size of the sample and pattern of the data, the AUC can be estimated with greater precision. None of the 95% confidence intervals fall below the fair range.

Table 31. AUC Values for the VPRAI-R, by Outcome of Interest

Outcome	AUC	CI (95%)
FTA	0.671	0.654-0.688
New Arrest	0.717	0.702-0.733
New Filing	0.667	0.649-0.685
New Conviction	0.655	0.632-0.677
New Violent Arrest	0.577	0.554-0.599
FTA or New Arrest	0.721	0.705-0.736

N = 3989

[32] See footnote 3.

Figure 14 shows the rate of various adverse outcomes during the pretrial period at each risk level of the VPRAI-R.[33] For each outcome of interest,[34] the observed rate of the outcome increases as the assessed risk level increases.

Figure 14. VPRAI-R Outcomes by Risk Category

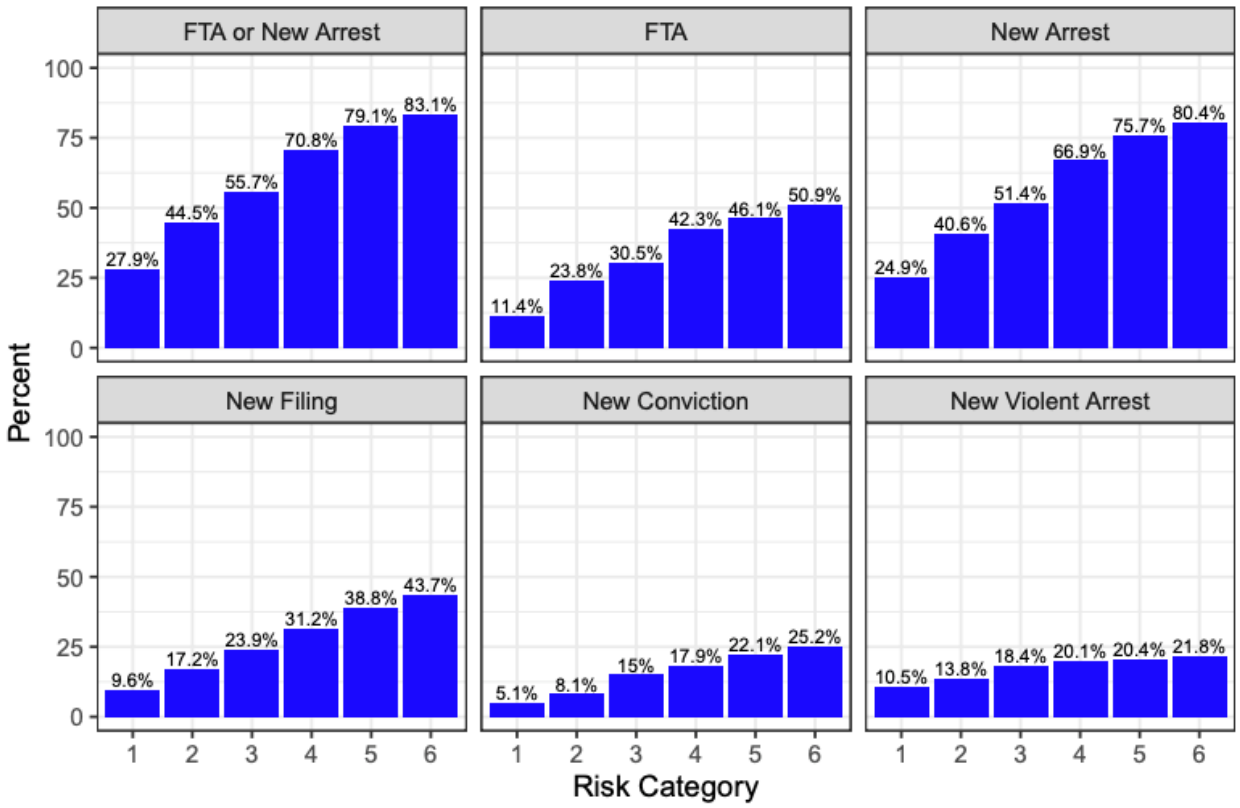


Table 32 shows the results from logistic regression models predicting each outcome of interest. The models control for the number of days the defendant spent released during the pretrial period. For each outcome of interest, the models show that the association between VPRAI-R risk score and the likelihood of the outcome occurring during the pretrial period is statistically significant ($p<0.001$).

[33] Risk levels are groupings of scores as defined by the tool developer.
[34] See validation methodology section for definitions of each outcome of interest.

Table 32. Logistic Regression Models Predicting the Likelihood of Outcomes of Interest by VPRAI-R Risk Scores Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
VPRAIR Risk Score	0.434***	0.565***	0.384***	0.350***	0.184***	0.588***
	(0.024)	(0.024)	(0.024)	(0.029)	(0.027)	(0.025)
Days Released	0.004***	0.003***	0.001***	0.001***	0.003***	0.004***
	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)
Constant	−3.001***	−2.236***	−2.575***	−3.221***	−2.761***	−2.228***
	(0.114)	(0.102)	(0.112)	(0.140)	(0.125)	(0.104)
Observations	3,989	3,989	3,989	3,989	3,989	3,989
Log Likelihood	−2,227.519	−2,360.564	−2,137.178	−1,582.911	−1,758.042	−2,294.661
Akaike Inf. Crit.	4,461.038	4,727.129	4,280.355	3,171.822	3,522.085	4,595.322
Note:						*p<0.05; **p<0.01; ***p<.001

ANALYSIS OF PREDICTIVE BIAS

RACE/ETHNICITY

Figure 15 shows the distribution of VPRAI-R risk assessment scores by race/ethnicity. The distribution of scores varies by race/ethnicity, with Hispanic individuals assessed more frequently in the lowest categories of risk while Black and White individuals mostly received scores in the midrange of the risk tool. Table 33 shows the counts associated with the risk distribution from Figure 15. The number of assessed individuals in each race/ethnicity group is sufficient to run statistical tests that look at how the VPRAIR-R tool scales performed by race/ethnicity.

Figure 15. Distribution of VPRAI-R Risk Scores by Race/Ethnicity

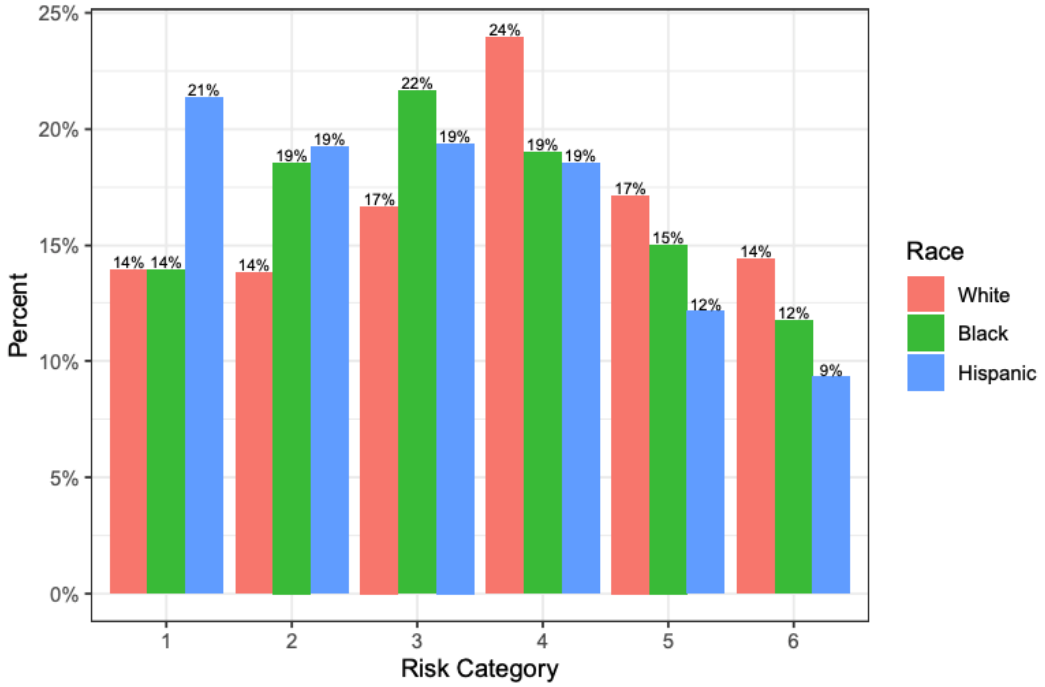


Table 33. Count of Individuals by VPRAI-R Risk Scores and Race/Ethnicity

VPRAIR Risk Score	White	Black	Hispanic
1	117	182	303
2	116	242	273
3	140	282	275
4	201	248	263
5	144	196	172
6	121	153	132

Table 34 shows the AUC values[35] and 95% confidence intervals for each outcome of interest for each race/ethnicity group. Most of the AUC values fall in the good to excellent range. However, AUCs for all race/ethnic groups fall into the fair range for new violent arrest, as do the AUC values for the Black group for the outcome FTA, new filing, and new conviction. Statistical testing[36] indicates a statistically significant difference between the AUC values for Black and Hispanic groups for FTA, new arrest, new violent arrest, and the combined measure of "FTA or New Arrest." For these outcomes, the AUC values for Black individuals are lower than the AUC values for Hispanic individuals, indicating that the VPRAI-R scale has a stronger ability to distinguish between individuals who are lower or higher risk for Hispanic

[35] See General Validation for a description of the meaning of AUC values.
[36] See Appendix D for DeLong's test for two ROC curves.

individuals than Black individuals. In addition, there is a statistically significant difference between the AUC values for the Black and White individuals for new arrest, new filing, new conviction, and the combined measure of "FTA or New Arrest." The AUC for Black individuals is lower than the AUC for White individuals in each of these outcomes, indicating that the VPRAI-R scale has a stronger ability to distinguish between individuals who are lower or higher risk for White individuals than Black individuals.

Table 34. AUC Values for the VPRAI-R, by Outcome of Interest and Race/Ethnicity

	AUC			CI (95%)		
Outcome	White	Black	Hispanic	White	Black	Hispanic
FTA	0.665	0.625	0.703	0.629-0.701	0.594-0.656	0.675-0.732
New Arrest	0.719	0.664	0.738	0.684-0.753	0.635-0.693	0.713-0.764
New Filing	0.689	0.631	0.667	0.652-0.726	0.598-0.664	0.635-0.699
New Conviction	0.680	0.615	0.647	0.637-0.722	0.574-0.656	0.607-0.687
New Violent Arrest	0.578	0.558	0.612	0.527-0.629	0.52-0.595	0.573-0.65
FTA or New Arrest	0.726	0.656	0.749	0.691-0.761	0.626-0.685	0.724-0.774

N White = 839 , N Black = 1303 , N Hispanic = 1418

Figure 16 shows the results of statistical models of the predictive power of the tool for each race/ethnicity group. Each line represents the probability of each outcome of interest at each risk level separately for each race/ethnicity. The grey area around each line represents a 95% confidence interval. Where the grey areas do not overlap, the evidence indicates there is likely a true difference between the groups; where the grey areas do overlap, the evidence is not strong enough to conclude that there are differences between them.

For the combined measure of "FTA or New Arrest," FTA, new arrest, new filing, and new conviction, the confidence intervals for all groups appear to be largely overlapping. This indicates that there is not strong evidence to conclude that the predictive power of the model differs markedly by race/ethnic group.

The confidence intervals are notably wider for the new violent arrest outcome because new violent arrest is a rarer outcome, which diminishes the ability of the model to make precise predictions. Nevertheless, the lack of overlap between the confidence intervals in the midrange of the tool indicates there is evidence that, for risk levels 2–4, Black individuals receiving this score have a higher probability of new violent arrest than White or Hispanic individuals with the same score.

Figure 16. Comparison of Racial/Ethnic Differences in Logistic Regression Curves by VPRAI-R Outcome of Interest

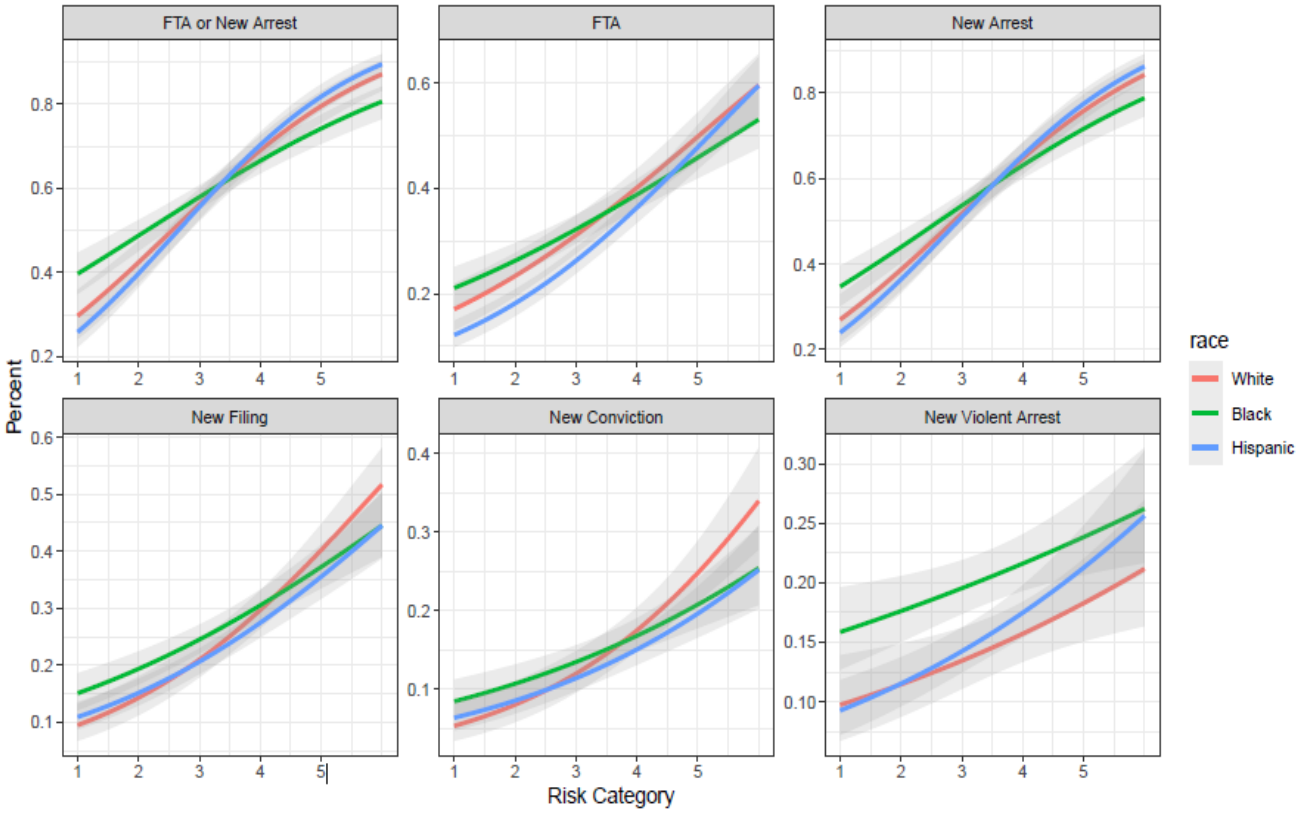


Table 35 shows the results of a logistic regression that predicts each outcome of interest by VPRAI-R risk score, race/ethnicity, and number of days spent released. Risk score is a statistically significant ($p<0.001$) predictor of the outcome of interest. The number of days an individual was out on release ($p<0.001$) also is a statistically significant predictor of all outcomes of interest, indicating that the longer an individual spends on release, the more likely the individual is to experience these outcomes of interest.

This statistical test compares Black and Hispanic individuals with White individuals. Black race was a statistically significant predictor ($p<0.01$) of new violent arrest. The positive coefficient indicates a higher probability of new violent arrest for Black individuals as compared to White individuals with the same risk score. Hispanic ethnicity is a statistically significant ($p<0.05$) predictor of FTA. The negative coefficient indicates a lower probability of FTA for Hispanic individuals as compared to White individuals with the same risk score.

This statistical test is limited, however, because it tests for an overall effect of race across the full risk scale, and, as can be seen from Figure 16, there may be some differences that emerge only in particular ranges of the tool. Table 36 shows the results of a more complex statistical model that allows for this possibility.

Table 35. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by VPRAI-R Risk Scores and Race/Ethnicity, Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
VPRAI-R Risk Score	0.433***	0.549***	0.375***	0.335***	0.204***	0.573***
	(0.025)	(0.026)	(0.026)	(0.031)	(0.029)	(0.027)
Race:Black	−0.071	−0.013	0.041	−0.075	0.391**	−0.018
	(0.099)	(0.099)	(0.102)	(0.122)	(0.121)	(0.101)
Race:Hispanic	−0.205*	−0.036	−0.113	−0.196	0.109	−0.037
	(0.100)	(0.097)	(0.104)	(0.125)	(0.125)	(0.099)
Days Released	0.004***	0.003***	0.001***	0.001***	0.003***	0.004***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Constant	−2.882***	−2.192***	−2.563***	−3.114***	−3.040***	−2.184***
	(0.142)	(0.132)	(0.143)	(0.176)	(0.168)	(0.135)
Observations	3,560	3,560	3,560	3,560	3,560	3,560
Log Likelihood	−2,001.384	−2,113.314	−1,911.235	−1,417.413	−1,556.942	−2,048.058
Akaike Inf. Crit.	4,012.768	4,236.627	3,832.470	2,844.826	3,123.884	4,106.116
Note:						*p<0.05; **p<0.01; ***p<.001

Table 36 shows the results of a logistic regression that predicts each outcome of interest by VPRAI-R risk score, race/ethnicity, the interaction between race and the VPRAI-R risk score, and number of days spent released. The number of days spent released is a statistically significant predictor of each outcome of interest. This statistical test again compares Black and Hispanic individuals with White individuals.

The results indicate that there is a statistically significant ($p<0.05$) interaction between Black race and VPRAI-R risk score on new arrest, new filing, new conviction, and the combined measure of "FTA or New Arrest" indicating that the effect of Black race on these outcomes of interest varies at different risk scores. The results indicate a statistically significantly higher probability of new filing and the combined outcome of "FTA or New Arrest" for Black individuals at the low end of the risk scale as compared to White individuals with the same risk score. At the high end of the scale, Black individuals have a significantly lower probability of a new conviction relative to White individuals with the same score.

There are no statistically significant interactions between VPRAI-R risk score and Hispanic ethnicity. In the absence of statistically significant interactions, the above model (Table 35) with no interaction is more appropriate to demonstrate the impact of Hispanic ethnicity.

Table 36. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by VPRAI-R Risk Scores, Race/Ethnicity, and Interaction of Race/Ethnicity and Risk Scores, Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
VPRAI-R Risk Score	0.425***	0.570***	0.466***	0.443***	0.189**	0.602***
	(0.051)	(0.053)	(0.055)	(0.066)	(0.062)	(0.054)
Race:Black	0.116	0.281	0.483*	0.440	0.458	0.363*
	(0.203)	(0.184)	(0.229)	(0.287)	(0.246)	(0.184)
Race:Hispanic	−0.437*	−0.170	0.148	0.176	−0.071	−0.208
	(0.204)	(0.180)	(0.227)	(0.284)	(0.249)	(0.180)
Days Released	0.004***	0.003***	0.001***	0.001***	0.003***	0.004***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
VPRAI-R*Black	−0.072	−0.129*	−0.152*	−0.167*	−0.028	−0.173*
	(0.065)	(0.066)	(0.069)	(0.083)	(0.077)	(0.068)
VPRAI-R*Hispanic	0.093	0.072	−0.085	−0.118	0.070	0.098
	(0.065)	(0.067)	(0.069)	(0.083)	(0.078)	(0.069)
Constant	−2.424***	−1.692***	−2.460***	−3.118***	−2.789***	−1.679***
	(0.175)	(0.159)	(0.196)	(0.244)	(0.217)	(0.160)
Observations	3,560	3,560	3,560	3,560	3,560	3,560
Log Likelihood	−1,997.229	−2,107.100	−1,908.777	−1,415.333	−1,555.756	−2,037.472
Akaike Inf. Crit.	4,008.457	4,228.200	3,831.554	2,844.666	3,125.512	4,088.944
Note:						*p<0.05; **p<0.01; ***p<.001

GENDER

Figure 17 shows the distribution of risk assessment scores by gender. The distribution of risk scores for women as compared to men is skewed toward the lower risk scores. Table 37 shows the counts associated with the risk distributions from Figure 17. The number of assessed individuals in each gender group is sufficient to run statistical tests that look at how the VPRAI-R tool performed by gender.

Figure 17. Distribution of VPRAI-R Risk Level by Gender

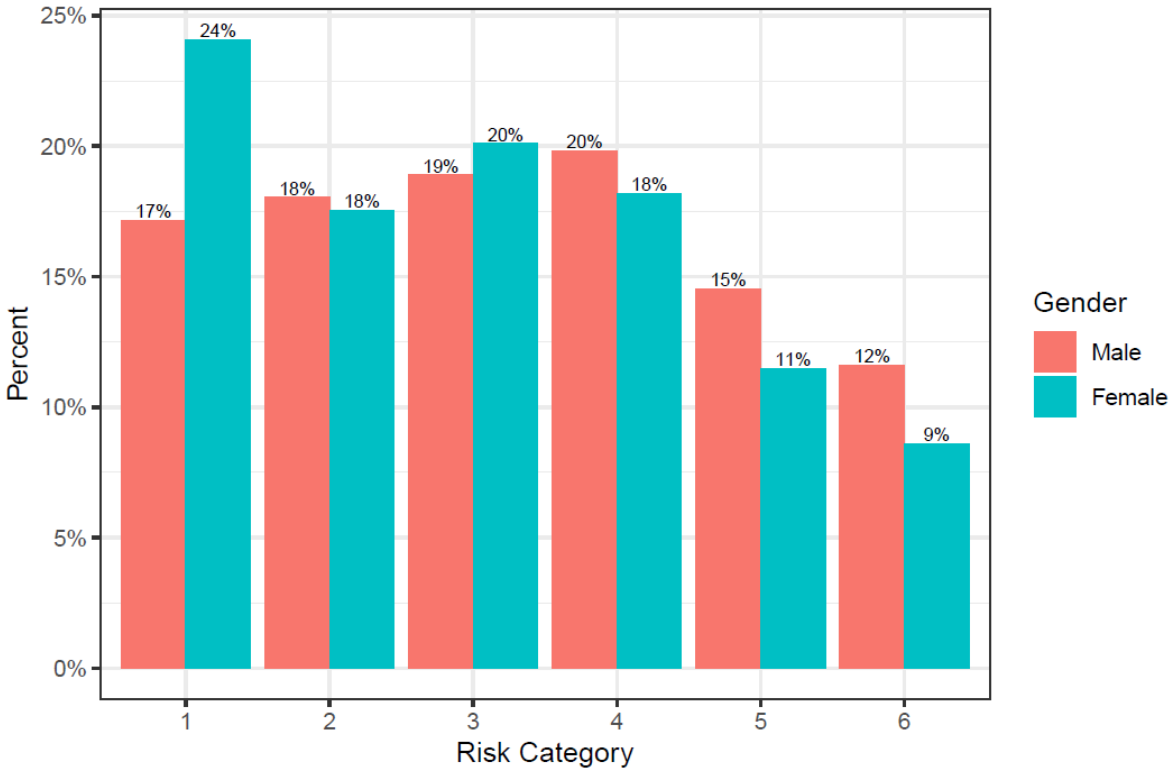


Table 37. Count of Individuals by VPRAI-R Risk Score and Gender

VPRAIR Risk Score	Male	Female
1	576	151
2	606	110
3	635	126
4	666	114
5	488	72
6	390	54

Table 38 shows the AUC values[37] and 95% confidence intervals for each outcome of interest and VPRAI-R risk score, separately for women and men. Apart from the new violent arrest outcome for men and women, and the new conviction AUC for women, which fall into the fair range, all other AUC values are

[37] See General Validation for a description of the meaning of AUC values.

in the good to excellent range. Statistical testing[38] indicates a statistically significant difference between female and male AUCs for new arrest and the combined outcome "FTA or New Arrest." The female AUC is lower than the male AUC for both of these outcomes, indicating that the VPRAI-R scale has a stronger ability to distinguish between individuals who are lower or higher risk of new arrest and "FTA or New Arrest" for men than women.

Table 38. AUC Values for Outcomes of Interest by Gender

	AUC		CI (95%)	
Outcome	Female	Male	Female	Male
FTA	0.646	0.676	0.602-0.69	0.658-0.694
New Arrest	0.668	0.726	0.626-0.709	0.709-0.743
New Filing	0.639	0.670	0.586-0.691	0.65-0.689
New Conviction	0.617	0.659	0.55-0.683	0.635-0.682
New Violent Arrest	0.580	0.575	0.521-0.639	0.551-0.599
FTA or New Arrest	0.672	0.729	0.63-0.714	0.712-0.746

N Female = 627 , N Male = 3361

Figure 18 shows the results of statistical models of the predictive power of the tool for each outcome of interest for women as compared to men. Each line represents the probability of each outcome of interest at each risk level separately for each race/ethnicity. The grey area around each line represents a 95% confidence interval: where the grey areas do not overlap, the evidence indicates there is likely a true difference between the groups; where the grey areas do overlap, the evidence is not strong enough to conclude that there are differences between them.

For new arrest and the combined measure of "FTA or New Arrest," the confidence intervals for women and men do not overlap at the higher ranges of the tool. This indicates that men have a higher likelihood of new arrest and the combined measure of "FTA or New Arrest" than women with the same score at the higher ranges of the scale. The confidence intervals are notably wider for the new violent arrest outcome because new violent arrest is a rarer outcome, which diminishes the ability of the model to make precise predictions.

[38] See Appendix D for DeLong's test for two ROC curves.

Figure 18. Comparison of Gender Differences in Logistic Regression Curves by VPRAI-R Outcome of Interest

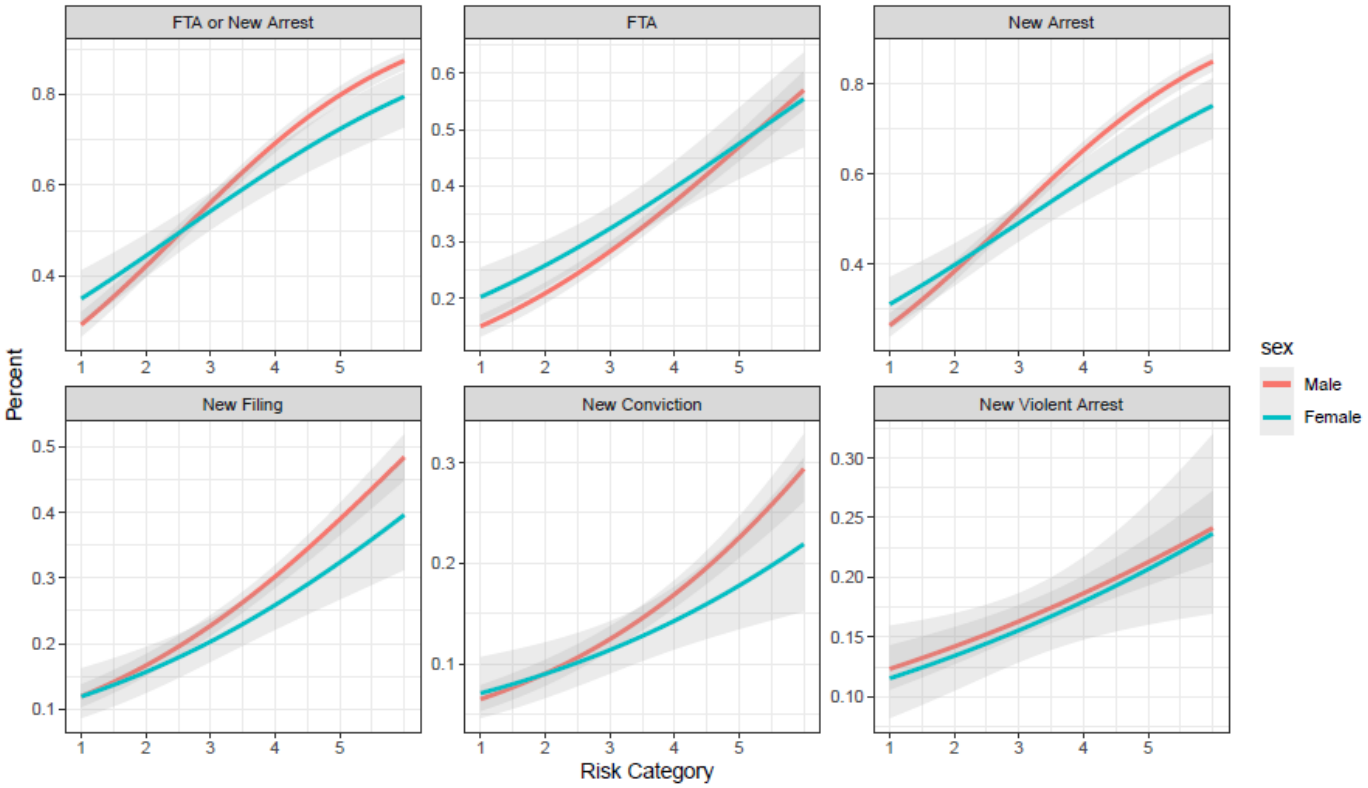


Table 39 shows the results of a logistic regression that predicts each outcome of interest by VPRAI-R risk score, gender, and number of days spent released. Risk score is a statistically significant ($p<0.001$) predictor of the outcome of interest. The number of days an individual was out on release ($p<0.001$) also is a statistically significant predictor of all outcomes of interest, indicating that the longer an individual spends on release, the more likely the individual is to experience these outcomes of interest. Results indicate that gender was not a statistically significant predictor of any outcome of interest.

This statistical test is limited, however, because it tests for an overall effect of race across the full risk scale, and, as can be seen from Figure 18, there may be some differences that emerge only in particular ranges of the tool. Table 40 shows the results of a more complex statistical model that allows for this possibility.

Table 39. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by VPRAI-R Risk Scores and Gender, Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
VPRAI-R Risk Score	0.436***	0.563***	0.381***	0.348***	0.183***	0.586***
	(0.024)	(0.024)	(0.024)	(0.029)	(0.027)	(0.025)
Female	0.154	−0.140	−0.187	−0.169	−0.067	−0.088
	(0.100)	(0.096)	(0.108)	(0.134)	(0.121)	(0.097)
Days Released	0.004***	0.003***	0.001***	0.001***	0.003***	0.004***
	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)
Constant	−3.033***	−2.208***	−2.540***	−3.189***	−2.747***	−2.210***
	(0.116)	(0.104)	(0.114)	(0.142)	(0.127)	(0.105)
Observations	3,988	3,988	3,988	3,988	3,988	3,988
Log Likelihood	−2,226.096	−2,358.997	−2,135.477	−1,581.997	−1,757.733	−2,293.669
Akaike Inf. Crit.	4,460.191	4,725.993	4,278.953	3,171.993	3,523.466	4,595.338
Note:						*p<0.05; **p<0.01; ***p<.001

Table 40 shows the results of a logistic regression that predicts each outcome of interest by VPRAI-R risk score, gender, the interaction between gender and the VPRAI-R risk score, and number of days spent released. The number of days spent released is a statistically significant predictor ($p<0.001$) of each outcome of interest. This statistical test again compares female with male individuals.

The results indicate that there is a statistically significant interaction between gender and VPRAI-R risk score on new arrest ($p<0.01$) and the combined measure of "FTA or New Arrest" ($p<0.05$), indicating that the effect of gender on these outcomes of interest varies at different risk scores. At the lower end of the scale, the differences are not significant. But at the higher end of the risk scale for new arrest and "FTA or New Arrest," females have a statistically significantly lower probability ($p<0.05$) of both new arrest and "FTA or New Arrest" when compared with males with the same score.

Table 40. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest, by VPRAI-R Risk Scores, and Interaction of Gender and Risk Scores, Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
VPRAI-R Risk Score	0.449***	0.592***	0.392***	0.362***	0.180***	0.614***
	(0.026)	(0.027)	(0.026)	(0.032)	(0.029)	(0.027)
Female	0.338	0.195	−0.012	0.079	−0.109	0.224
	(0.175)	(0.156)	(0.200)	(0.252)	(0.212)	(0.155)
Days Released	0.004***	0.003***	0.001***	0.001***	0.003***	0.004***
	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)
VPRAI-R*Female	−0.079	−0.167**	−0.069	−0.094	0.018	−0.163*
	(0.062)	(0.062)	(0.067)	(0.083)	(0.074)	(0.063)
Constant	−2.631***	−1.705***	−2.186***	−2.880***	−2.557***	−1.679***
	(0.101)	(0.088)	(0.099)	(0.124)	(0.110)	(0.089)
Observations	3,988	3,988	3,988	3,988	3,988	3,988
Log Likelihood	−2,225.301	−2,355.443	−2,134.957	−1,581.354	−1,757.703	−2,290.456
Akaike Inf. Crit.	4,460.603	4,720.887	4,279.914	3,172.707	3,525.406	4,590.913

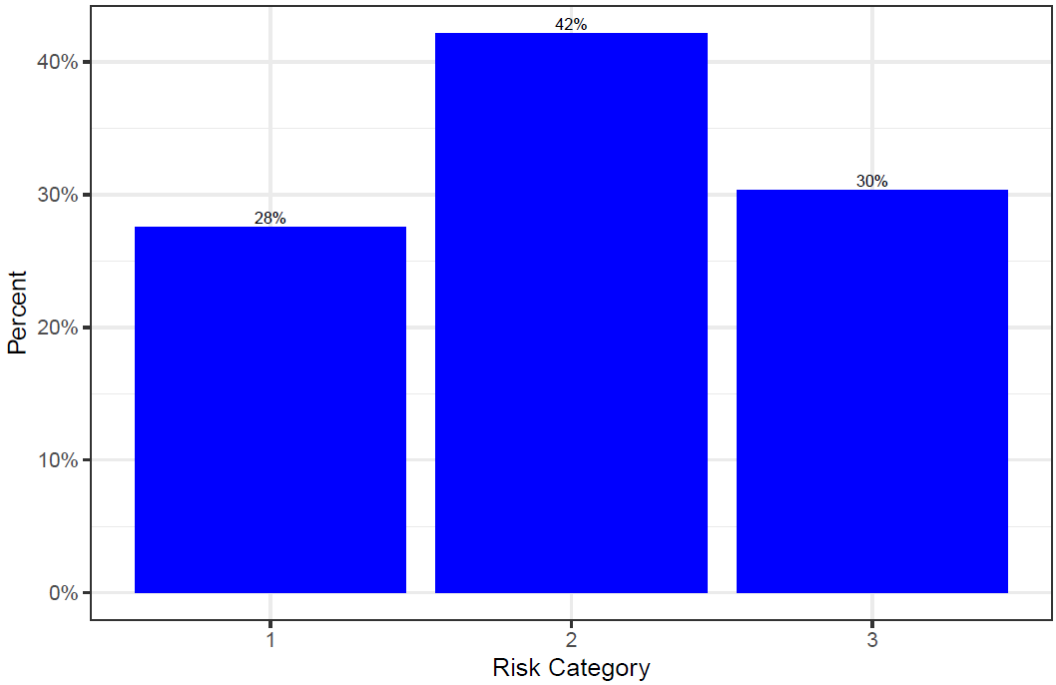
Note: *p<0.05; **p<0.01; ***p<.001

OHIO RISK ASSESSMENT: PRETRIAL ASSESSMENT TOOL (ORAS-PAT) VALIDATION

GENERAL VALIDATION

Figure 19 shows the distribution of risk categories for individuals in the evaluation data set assessed with the ORAS tool, used by the Pretrial Pilot Projects in Modoc, Napa, Nevada, Ventura, and Yuba Counties. The ORAS tool developer divided the risk scores into three risk levels: level 1 includes scores 0–2; level 2, scores 3–5; and level 3, scores 6–9.[39] Risk level 2 (scores 3–5) was the most common assessed risk level in the evaluation data set, followed by level 3 and level 1. The distribution of all assessed individuals may differ from the distribution in the evaluation data set because the evaluation data set includes only released individuals with concluded pretrial periods. Table 41 shows the counts associated with the distribution in Figure 19.

Figure 19. Distribution of ORAS Risk Scores



[39] Edward J. Latessa, Richard Lemke, Matthew Makarios, Paul Smith, and Christopher T. Lowenkamp, "The Creation and Validation of the Ohio Risk Assessment System (ORAS)" (2010) 74(1) *Fed. Probation* 16.

Table 41. Counts of Individuals by ORAS Risk Scores

ORAS Risk Level	Total
1	895
2	1,369
3	982

Table 42 shows the AUC value for the ORAS tool, using the three established risk levels for each outcome of interest. The AUC value is a single number that represents the ability of the tool to discriminate between individuals who are lower or higher risk across the range of the tool. For criminal justice risk assessments, a common metric for evaluating AUC values is derived from Desmarais and Singh (2013),[40] who defined AUC values less than 0.55 as poor, 0.55–0.63 as fair, 0.64–0.70 as good, and 0.71–1.00 as excellent. By these definitions, the AUC values for the ORAS are excellent for new arrest, new conviction, and the combined measure of "FTA or New Arrest"; good for FTA and new filing; and fair for new violent arrest.

The 95% confidence interval is also shown. It represents the range of AUC estimates that the true AUC value is 95% likely to fall between. A smaller range indicates that given the size of the sample and pattern of the data, the AUC can be estimated with greater precision. None of the 95% confidence intervals fall below the fair range.

Table 42. AUC Values for the ORAS, by Outcome of Interest

Outcome	AUC	CI (95%)
FTA	0.661	0.64-0.681
New Arrest	0.724	0.708-0.741
New Filing	0.701	0.679-0.724
New Conviction	0.706	0.682-0.731
New Violent Arrest	0.634	0.605-0.663
FTA or New Arrest	0.715	0.698-0.731

N = 3246

Figure 20 shows the rate of various adverse outcomes during the pretrial period at each risk level of the ORAS.[41] For each outcome of interest,[42] observed rates of the outcome increase consistently as the assessed risk level increases. The ORAS tool was specifically designed to predict a combination of risk of failure to appear in court and risk of a new arrest.

[40] See note 3.

[41] Risk levels are groupings of scores as defined by the tool developer.

[42] See the validation methodology section for definitions of each outcome of interest.

Figure 20. ORAS Outcomes by Risk Category

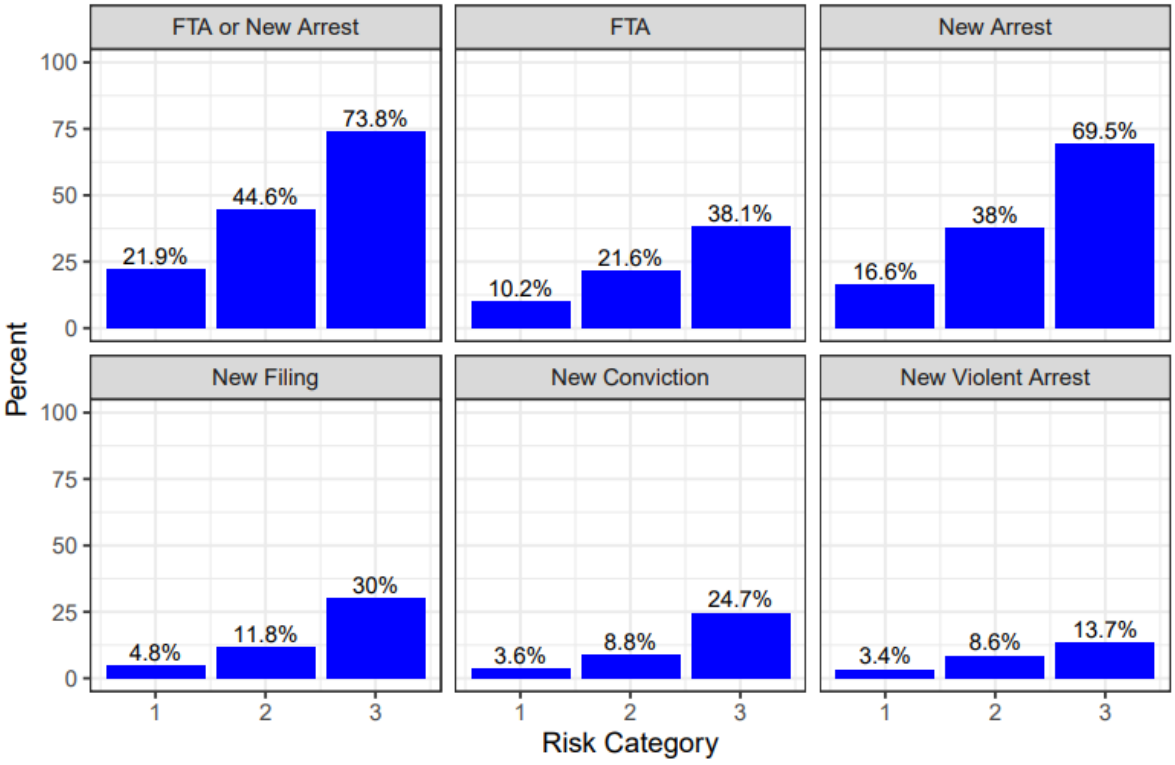


Table 43 shows the results from logistic regression models predicting each outcome of interest. The models control for the number of days the defendant spent released during the pretrial period. For each outcome of interest, the models show that the ORAS risk score is statistically significantly ($p<0.001$) associated with the likelihood of all outcomes during the pretrial period, as is the number of days an individual spends on release.

Table 43. Logistic Regression Models Predicting the Likelihood of Outcomes of Interest by ORAS Risk Scores Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
ORAS Risk Score	0.831***	1.224***	1.086***	1.117***	0.643***	1.171***
	(0.063)	(0.057)	(0.079)	(0.088)	(0.090)	(0.056)
Days Released	0.004***	0.002***	0.003***	0.003***	0.003***	0.003***
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0002)
Constant	−3.764***	−3.276***	−4.829***	−5.189***	−4.288***	−3.015***
	(0.163)	(0.136)	(0.210)	(0.236)	(0.228)	(0.132)
Observations	3,246	3,246	3,246	3,246	3,246	3,246
Log Likelihood	−1,511.106	−1,864.858	−1,185.416	−1,026.305	−893.208	−1,885.030
Akaike Inf. Crit.	3,028.212	3,735.716	2,376.833	2,058.609	1,792.416	3,776.060
Note:						*p<0.05; **p<0.01; ***p<.001

ANALYSIS OF PREDICTIVE BIAS

RACE/ETHNICITY

Figure 21 shows the distribution of ORAS risk assessment scores by race/ethnicity. The distribution of scores varies by race/ethnicity, with Black individuals assessed more frequently in the midrange of the tool and less frequently in the high range of the tool relative to White and Hispanic individuals. Table 44 shows the counts associated with the risk distributions from Figure 21. The number of assessed individuals in each race/ethnicity group (Table 44) is sufficient to run statistical tests that look at how the VPRAIR-R tool scales performed by race/ethnicity.

Figure 21. Distribution of ORAS Risk Scores by Race/Ethnicity

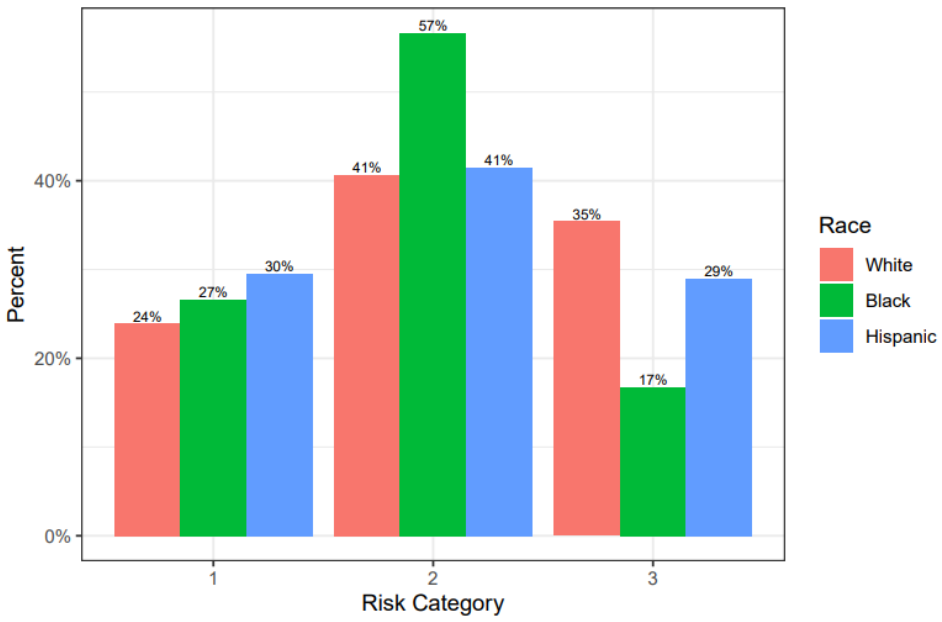


Table 44. Count of Individuals by ORAS Risk Scores and Race/Ethnicity

ORAS Risk Score	White	Black	Hispanic
1	285	54	516
2	484	115	724
3	422	34	506

Table 45 shows the AUC values[43] and 95% confidence intervals for each outcome of interest for each race/ethnicity group. The AUC values for new violent arrest fall into the fair range. All other AUC values fall in the good to excellent range. Statistical testing[44] indicates that there are no statistically significant differences in AUC values among the Black, Hispanic, and White groups.

[43] See General Validation for a description of the meaning of AUC values.
[44] See Appendix D for DeLong's test for two ROC curves.

Table 45. AUC Values for the ORAS, by Outcome of Interest and Race/Ethnicity

	AUC			CI (95%)		
Outcome	White	Black	Hispanic	White	Black	Hispanic
FTA	0.660	0.660	0.658	0.628-0.692	0.582-0.738	0.631-0.686
New Arrest	0.724	0.661	0.720	0.698-0.751	0.589-0.732	0.697-0.742
New Filing	0.702	0.696	0.687	0.669-0.736	0.6-0.791	0.655-0.72
New Conviction	0.708	0.710	0.694	0.672-0.745	0.592-0.829	0.659-0.729
New Violent Arrest	0.631	0.609	0.631	0.584-0.678	0.463-0.755	0.592-0.671
FTA or New Arrest	0.725	0.645	0.706	0.699-0.752	0.576-0.713	0.683-0.728

N White = 1191 , N Black = 203 , N Hispanic = 1746

Figure 22 shows the results of statistical models of the predictive power of the relevant ORAS scale for each outcome of interest for each race/ethnicity group. Each line represents the probability of each outcome of interest at each risk score separately for each race/ethnicity group. The grey area around each line represents a 95% confidence interval. When the grey areas do not overlap, the evidence indicates that there is likely a true difference between the groups. Conversely, when the grey areas do overlap, the evidence may not be strong enough to conclude that there are differences between them.

For all outcomes, the confidence intervals of the lines for the Black, White, and Hispanic race/ethnicity groups overlap. Thus, there may be insufficient evidence to conclude any true difference in the likelihood of those outcomes since individuals differ across these groups with the same score. The 95% confidence intervals are notably wider for new violent arrest due to the small sample size, which diminishes the ability of the models to make precise predictions for this specific outcome.

Figure 22. Comparison of Racial/Ethnic Differences in Logistic Regression Curves by ORAS Outcome of Interest

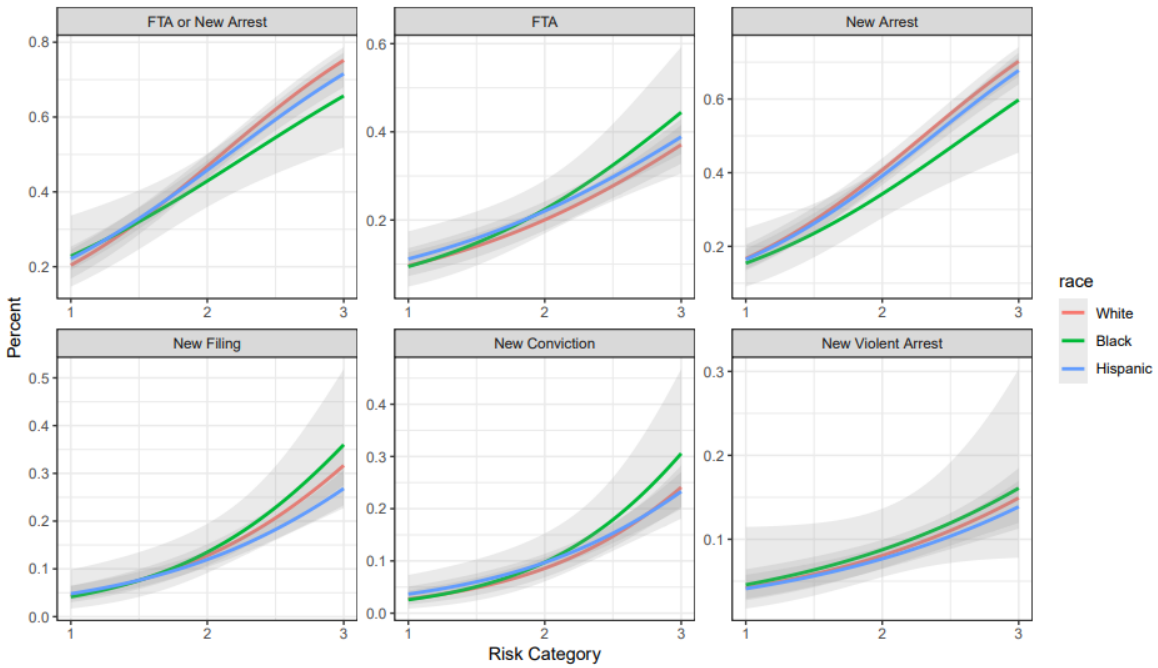


Table 46 shows the results of a logistic regression, which predicts each outcome of interest by the relevant ORAS risk score, race, and number of days spent released. This statistical test compares Black and Hispanic individuals with White individuals. The ORAS risk score is a statistically significant ($p<0.001$) predictor of all outcomes of interest. The number of days an individual was out on release is also a statistically significant predictor ($p<0.001$) of all outcomes of interest, indicating that the longer an individual spends on release, the more likely the individual is to experience these outcomes.

Black race is a statistically significant ($p<0.05$) predictor for new arrest. The negative coefficient on Black race indicates that Black individuals have a lower probability of a new arrest as compared to White individuals with the same risk score. Hispanic ethnicity was not a statistically significant predictor for any outcome of interest.

This statistical test is limited, however, because it tests for an overall effect of race across the full risk scale. As can be seen from the above charts, there may be different patterns across particular ranges of the tool. Table 47 shows the results of a more complex statistical model that allows for this possibility.

Table 46. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by ORAS Risk Scores and Race/Ethnicity, Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
ORAS Risk Score	0.826***	1.189***	1.045***	1.088***	0.621***	1.144***
	(0.065)	(0.058)	(0.080)	(0.090)	(0.091)	(0.057)
Race:Black	0.069	−0.347*	−0.027	0.072	0.014	−0.227
	(0.200)	(0.175)	(0.230)	(0.255)	(0.274)	(0.170)
Race:Hispanic	0.090	−0.093	−0.185	0.021	−0.089	−0.078
	(0.097)	(0.085)	(0.111)	(0.122)	(0.135)	(0.084)
Days Released	0.004***	0.002***	0.003***	0.003***	0.003***	0.003***
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0002)
Constant	−3.806***	−3.126***	−4.640***	−5.143***	−4.174***	−2.896***
	(0.180)	(0.149)	(0.224)	(0.256)	(0.248)	(0.145)
Observations	3,140	3,140	3,140	3,140	3,140	3,140
Log Likelihood	−1,466.023	−1,812.448	−1,152.165	−995.327	−876.283	−1,828.343
Akaike Inf. Crit.	2,942.046	3,634.895	2,314.330	2,000.654	1,762.566	3,666.687

Note: *p<0.05; **p<0.01; ***p<.001

Table 47 shows the results of a logistic regression model that predicts each outcome of interest by the relevant ORAS scale risk score, race, the interaction between race and the ORAS risk score, and number of days spent released. The number of days released is a statistically significant predictor (p<0.001) of each outcome of interest. This statistical test again compares Black and Hispanic individuals with White individuals. There are no statistically significant interactions between race/ethnicity and the relevant ORAS scale risk scores on any of the outcomes of interest. In the absence of a statistically significant interaction, the model shown in Table 46 with no interaction is more appropriate to demonstrate the impact of race/ethnicity.

Table 47. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by Risk Scores, Race/Ethnicity, and Interaction of Race/Ethnicity and Risk Scores, Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
ORAS Risk Score	0.844***	1.227***	1.133***	1.184***	0.640***	1.244***
	(0.106)	(0.093)	(0.129)	(0.151)	(0.146)	(0.093)
Race:Black	−0.041	−0.102	−0.082	−0.060	0.079	0.136
	(0.407)	(0.335)	(0.523)	(0.617)	(0.556)	(0.309)
Race:Hispanic	0.141	−0.037	0.051	0.282	−0.050	0.081
	(0.201)	(0.163)	(0.265)	(0.311)	(0.292)	(0.155)
Days Released	0.004***	0.002***	0.003***	0.003***	0.003***	0.003***
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0002)
ORAS*Black	0.105	−0.232	0.065	0.125	−0.053	−0.362
	(0.313)	(0.277)	(0.372)	(0.421)	(0.415)	(0.267)
ORAS*Hispanic	−0.040	−0.048	−0.165	−0.176	−0.029	−0.145
	(0.136)	(0.120)	(0.167)	(0.191)	(0.190)	(0.119)
Constant	−3.004***	−1.983***	−3.724***	−4.203***	−3.579***	−1.866***
	(0.172)	(0.135)	(0.219)	(0.264)	(0.239)	(0.131)
Observations	3,140	3,140	3,140	3,140	3,140	3,140
Log Likelihood	−1,465.887	−1,812.084	−1,151.572	−994.726	−876.267	−1,827.045
Akaike Inf. Crit.	2,945.774	3,638.168	2,317.143	2,003.453	1,766.535	3,668.090
Note:						*p<0.05; **p<0.01; ***p<.001

GENDER

Figure 23 shows the distribution of risk assessment scores by gender. The distribution of risk scores between men and women is similar. Table 48 shows the counts associated with the risk distributions from Figure 23. The number of assessed individuals in each gender group (Table 48) is sufficient to run statistical tests that examine how the ORAS tool performed by gender.

Figure 23. Distribution of ORAS Risk Scores by Gender

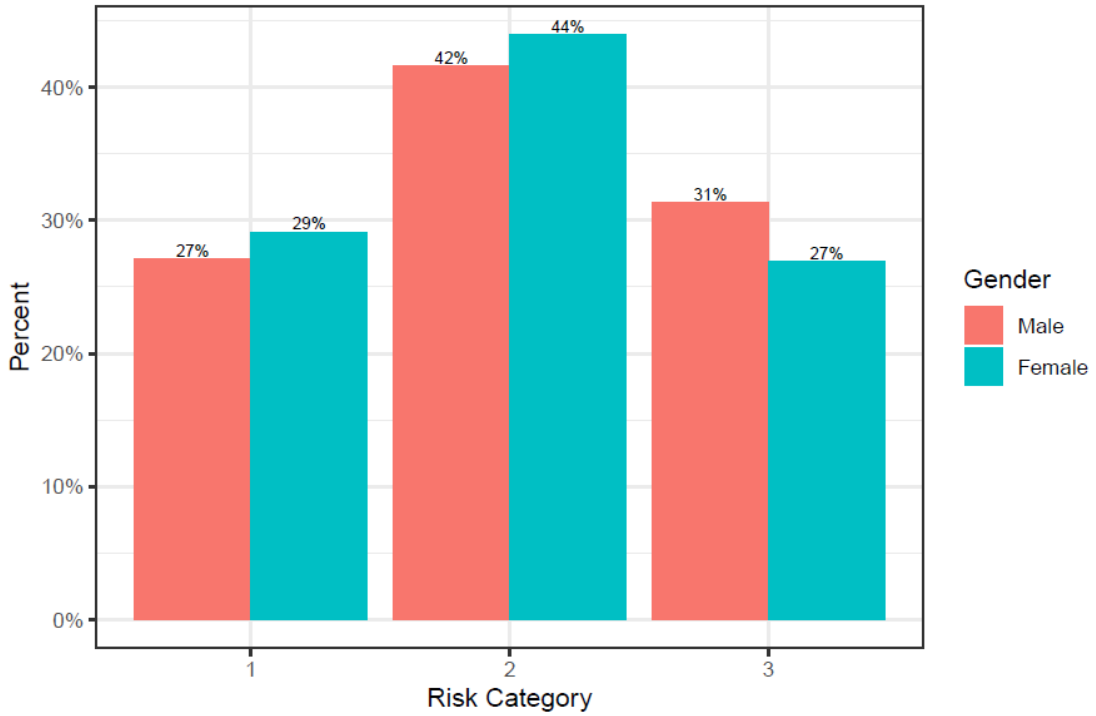


Table 48. Count of Individuals by ORAS Risk Scores and Gender

ORAS Risk Score	Male	Female
1	683	212
2	1,050	318
3	786	196

Table 49 shows the AUC values[45] and 95% confidence intervals for each outcome of interest and relevant ORAS risk scales separately for men and women. Apart from the new violent arrest AUC value for men, which falls in the fair range, all other AUC values are in the good to excellent range. Statistical testing[46] indicates that there are no statistically significant differences in AUC between men and women for any of the outcomes of interest.

[45] See General Validation for a description of the meaning of AUC values.
[46] See Appendix D for DeLong's test for two ROC curves.

Table 49. AUC Values for the ORAS, by Outcome of Interest and Gender

	AUC		CI (95%)	
Outcome	Female	Male	Female	Male
FTA	0.641	0.665	0.598-0.685	0.642-0.687
New Arrest	0.717	0.726	0.681-0.752	0.708-0.744
New Filing	0.698	0.701	0.644-0.752	0.676-0.726
New Conviction	0.697	0.707	0.633-0.761	0.68-0.733
New Violent Arrest	0.640	0.632	0.58-0.7	0.599-0.666
FTA or New Arrest	0.713	0.715	0.678-0.748	0.696-0.733

N Female = 726 , N Male = 2519

Figure 24 shows the results of statistical models of the predictive power of the relevant ORAS scale for each outcome of interest for women as compared to men. Each line represents the probability of each outcome of interest at each risk level separately by gender. The grey area around each line represents a 95% confidence interval: where the grey areas do not overlap, the evidence indicates there is likely a true difference between the groups; where the grey areas do overlap, the evidence may not be strong enough to conclude that there are differences between them.

For FTA and the composite measure “FTA or New Arrest,” the overlapping confidence intervals for men and women suggest that the probability of an FTA and “FTA or New Arrest” during the pretrial period is the same for men and women with the same risk score.

For the new arrest, new filing, and new conviction outcomes, there are areas where the confidence intervals do not overlap, which provides evidence that women had a lower likelihood of these outcomes during the pretrial period, particularly in the mid-to-high range of the scale as compared to men with the same risk score. For new violent arrest, the confidence intervals are notably wider since it is a rarer outcome, which diminishes the ability of the model to make precise predictions.

Figure 24. Comparison of Gender Difference in Logistic Regression Curves by ORAS Outcome of Interest

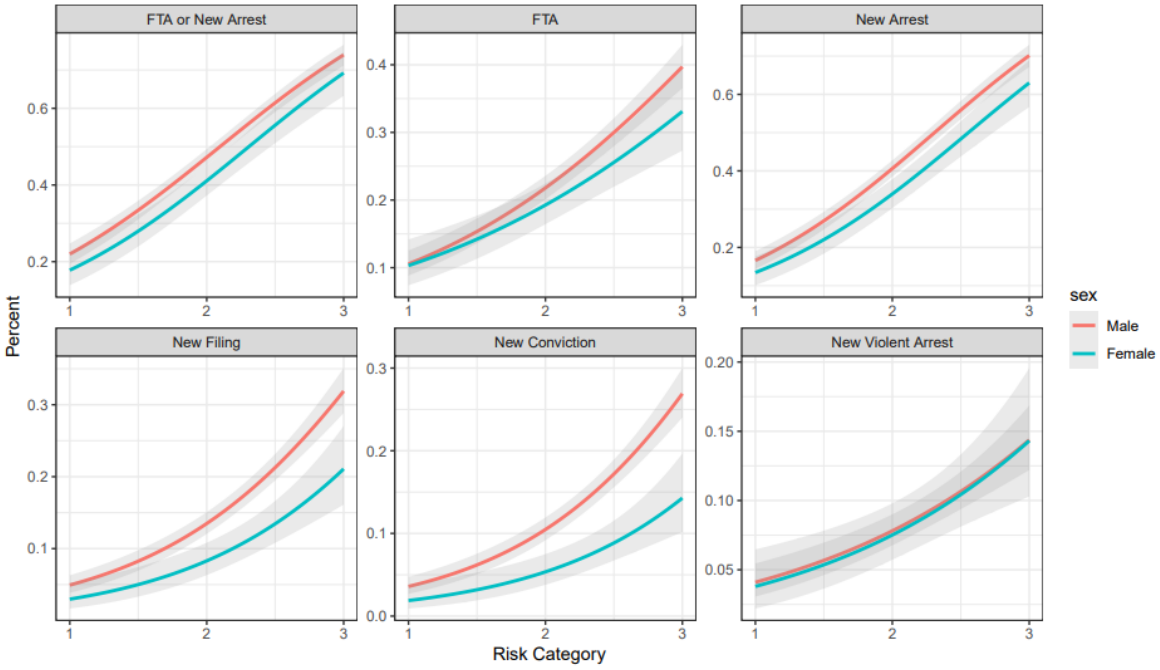


Table 50 shows the results of a logistic regression that predicts each outcome of interest by the ORAS risk score, gender, and number of days spent released. This statistical test compares women with the base group of men. Risk level on the ORAS is, in each case, a statistically significant ($p<0.001$) predictor of the outcome of interest. The number of days the individual was out on release was also a statistically significant predictor ($p<0.001$), indicating that the longer an individual spends on release, the more likely the individual is to experience the outcome of interest.

Female gender is a statistically significant predictor for the outcomes of new arrest ($p<0.01$), new filing ($p<0.001$), new conviction ($p<0.001$), and the composite measure of "FTA or New Arrest" ($p<0.05$). The negative coefficients indicate that women are statistically significantly less likely to experience the outcomes of new arrest, new filing, new conviction, and the combined measure of "FTA or New Arrest" as compared to men with the same risk score.

This statistical test is limited, however, because it tests for an overall effect of race across the full risk scale, and, as can be seen from the above charts, there may be different patterns across particular ranges of the tool. Table 51 presents the results of a more complex statistical model that allows for this possibility.

Table 50. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by ORAS Risk Scores and Gender, Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
ORAS Risk Score	0.829***	1.222***	1.079***	1.107***	0.644***	1.170***
	(0.063)	(0.057)	(0.079)	(0.088)	(0.090)	(0.056)
Female	−0.112	−0.256**	−0.506***	−0.720***	0.028	−0.204*
	(0.112)	(0.097)	(0.142)	(0.167)	(0.156)	(0.095)
Days Released	0.004***	0.002***	0.003***	0.003***	0.003***	0.003***
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0002)
Constant	−3.733***	−3.212***	−4.712***	−5.035***	−4.295***	−2.963***
	(0.165)	(0.138)	(0.211)	(0.238)	(0.233)	(0.133)
Observations	3,245	3,245	3,245	3,245	3,245	3,245
Log Likelihood	−1,510.478	−1,860.952	−1,178.584	−1,015.751	−893.141	−1,882.289
Akaike Inf. Crit.	3,028.957	3,729.903	2,365.168	2,039.503	1,794.282	3,772.579
Note:						*p<0.05; **p<0.01; ***p<.001

Table 51 shows the results of a logistic regression model that predicts each outcome of interest by the ORAS risk score, gender, the interaction between gender and the ORAS risk score, and number of days spent released. The results indicate that there is no statistically significant interaction between gender and risk score for any of the outcomes of interest. In the absence of statistically significant interactions, the model shown in Table 50 with no interaction is more appropriate to demonstrate the impact of gender.

Table 51. Logistic Regression Model Predicting the Likelihood of Outcomes of Interest by ORAS Risk Scores, Gender, and Interaction of Gender and Risk Scores, Controlling for Days Released

	Dependent variable:					
	FTA	New Arrest	New Filing	New Conviction	New Violent Arrest	FTA or New Arrest
	(1)	(2)	(3)	(4)	(5)	(6)
ORAS Risk Score	0.845***	1.225***	1.072***	1.108***	0.625***	1.160***
	(0.071)	(0.064)	(0.086)	(0.095)	(0.101)	(0.063)
Female	−0.012	−0.241	−0.564	−0.707	−0.086	−0.252
	(0.220)	(0.187)	(0.336)	(0.408)	(0.329)	(0.174)
Days Released	0.004***	0.002***	0.003***	0.003***	0.003***	0.003***
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0002)
ORAS*Female	−0.082	−0.013	0.042	−0.009	0.088	0.045
	(0.155)	(0.139)	(0.215)	(0.257)	(0.221)	(0.136)
Constant	−2.925***	−1.992***	−3.624***	−3.930***	−3.627***	−1.783***
	(0.119)	(0.096)	(0.150)	(0.168)	(0.168)	(0.092)
Observations	3,245	3,245	3,245	3,245	3,245	3,245
Log Likelihood	−1,510.340	−1,860.947	−1,178.565	−1,015.751	−893.062	−1,882.234
Akaike Inf. Crit.	3,030.681	3,731.894	2,367.130	2,041.502	1,796.124	3,774.467
Note:						*p<0.05; **p<0.01; ***p<.001

VPRAI-O VALIDATION

The sample size for the VPRAI-O was smaller than the designated minimum sample size of 200 for general validation. Therefore, a validation of the VPRAI-O risk assessment tool is not included in this report.

APPENDIX A

TABLE A1. PSA FACTORS AND WEIGHTS

Risk Factor	Response	Weight
FAILURE TO APPEAR (0–7)		
Pending Charge at the Time of the Offense	No	**0**
	Yes	**1**
Prior Conviction	No	**0**
	Yes	**1**
Prior Failures to Appear Pretrial in Past 2 Years	0	**0**
	1	**2**
	2 or More	**4**
Prior Failure to Appear Pretrial Older than 2 Years	No	**0**
	Yes	**1**
NEW CRIMINAL ACTIVITY (0–13)		
Age at Current Arrest	23 or Older	**0**
	22 or Younger	**2**
Pending Charge at the Time of the Offense	No	**0**
	Yes	**3**
Prior Misdemeanor Conviction	No	**0**
	Yes	**1**
Prior Felony Conviction	No	**0**
	Yes	**1**
Prior Violent Conviction	0	**0**
	1 or 2	**1**
	3 or more	**2**
Prior Failure to Appear Pretrial in Past 2 Years	0	**0**
	1	**1**
	2 or More	**2**
Prior Sentence to Incarceration	No	**0**
	Yes	**2**
NEW VIOLENT CRIMINAL ACTIVITY (0–7)		
Current Violent Offense	No	**0**
	Yes	**2**
Current Violent Offense & 20 Years Old or Younger	No	**0**
	Yes	**1**
Pending Charge at the Time of the Offense	No	**0**
	Yes	**1**
Prior Conviction	No	**0**
	Yes	**1**
Prior Violent Conviction	0	**0**
	1 or 2	**1**
	3 or More	**2**

Source: *Public Safety Assessment: Risk Factors and Formula*, Laura and John Arnold Foundation (2013)

Outcome Measure	Raw Score	Risk Scale
Failure to Appear (FTA)	0	**1**
	1	**2**
	2	**3**
	3	**4**
	4	**4**
	5	**5**
	6	**5**
	7	**6**
New Criminal Activity (NCA)	0	**1**
	1	**2**
	2	**2**
	3	**3**
	4	**3**
	5	**4**
	6	**4**
	7	**5**
	8	**5**
	9–13	**6**
Outcome Measure	**Raw Score**	**NCVA Flag**
New Violent Criminal Activity (NVCA)	0	**No**
	1	**No**
	2	**No**
	3	**No**
	4	**Yes**
	5	**Yes**
	6	**Yes**
	7	**Yes**

Source: *Public Safety Assessment: Risk Factors and Formula*, Laura and John Arnold Foundation (2013)

TABLE A2. VPRAI FACTORS AND WEIGHTS

Risk Factor	Criteria	Weight
Primary Charge Type	If the most serious charge for the current arrest was a felony	**No = 0**
		Yes = 1
Pending Charge(s)	If the defendant had one or more charge(s) pending in court at the time of the arrest	**No = 0**
		Yes = 1
Criminal History	If the defendant had one or more misdemeanor or felony convictions	**No = 0**
		Yes = 1
Two or More Failures to Appear	If the defendant had two or more failure to appear convictions	**No = 0**
		Yes = 2
Two or More Violent Convictions	If the defendant had two or more violent convictions	**No = 0**
		Yes = 1
Length at Current Residence	If the defendant had lived at their current residence for less than one year prior to arrest	**No = 0**
		Yes = 1
Employed/Primary Caregiver	If the defendant had not been employed continuously for the past two years and was not the primary caregiver for a child at the time of arrest	**No = 0**
		Yes = 1
History of Drug Abuse	If the defendant had a history of drug abuse	**No = 0**
		Yes = 1
Point Range		**0–9**

Source: *Pretrial Risk Assessment in Virginia: The Virginia Pretrial Risk Assessment Instrument*, Virginia Department of Criminal Justice (2009)

TABLE A3. VIRGINIA PRETRIAL RISK ASSESSMENT INSTRUMENT, *REVISED*: FACTORS AND WEIGHTS

Risk Factor	Criteria	Weight
Active Community Criminal Justice Supervision	If the defendant is under active community supervision at the time of arrest	**No = 0**
		Yes = 2
Charge is Felony Drug, Felony Theft, or Felony Fraud	If the defendant's charge is felony drug, felony theft, or felony fraud	**No = 0**
		Yes = 3
Pending Charge(s)	If the defendant had one or more charge(s) pending in court at the time of the arrest	**No = 0**
		Yes = 2
Criminal History	If the defendant had one or more misdemeanor or felony convictions	**No = 0**
		Yes = 2
Two or More Failures to Appear	If the defendant had two or more failure to appear convictions	**No = 0**
		Yes = 1
Two or More Violent Convictions	If the defendant had two or more violent convictions	**No = 0**
		Yes = 1
Unemployed at Time of Arrest	If the defendant is unemployed, a full-time student, a primary caregiver, or a retiree at the time of arrest	**No = 0**
		Yes = 1
History of Drug Abuse	If the defendant had a history of drug abuse	**No = 0**
		Yes = 2
Point Range		**0–14**

TABLE A4. ORAS-PAT FACTORS AND WEIGHTS

Risk Factor	Response	Weight
Age at First Arrest	33 or Older	**0**
	Under 33	**1**
Number of Failure-to-Appear Warrants in Past 24 Months	None	**0**
	One Warrant for FTA	**1**
	Two or More FTA Warrants	**2**
Three or More Prior Jail Incarcerations	No	**0**
	Yes	**1**
Employed at the Time of Arrest	Yes, Full-time	**0**
	Yes, Part-time	**1**
	Not Employed	**2**
Residential Stability	Lived at Current Residence for Past Six Months	**0**
	Not Lived at Same Residence	**1**
Illegal Drug Use During Past Six Months	No	**0**
	Yes	**1**
Severe Drug Use Program	No	**0**
	Yes	**1**
Point Range		**0–9**

Source: *Creation and Validation of the Ohio Risk Assessment System: Final Report*, University of Cincinnati School of Criminal Justice, Center for Criminal Justice Research (2009)

APPENDIX B

TABLE B1. PSA VIOLENT OFFENSE LIST

PENAL CODE	DESCRIPTION
69	Obstructing or resisting exec officer in performance of duty; threats, force, or violence
136.1(c)(1)	Intimidating/Threat Witness/Victim and Act is accompanied by force
140(a)	Threatening Witnesses, victims or informants
148(b)	Removal or taking of weapon other than firearm from peace officer during commission of resisting offense
148(c)	Removal or taking of firearm from peace officer during commission of resisting offense
148(d)	Removal or taking of weapon firearm from peace officer engaged in performance of duty
148.10(a)	Resist Po: Cause death/SBI
149	Assault by a public officer
151	Advocacy to kill or injure peace officer
186.26(c)	Use of coercion or violence to solicit or recruit another to actively participate in criminal street gang
187(a)	Murder first or second degree
191.5(a)	Gross vehicular manslaughter while intoxicated
192(a)	Voluntary manslaughter
192(b)	Involuntary manslaughter
192(c)(1)	Vehicular manslaughter with gross negligence
192(c)(3)	Vehicular manslaughter
192.5(a)	Vehicular manslaughter in the operation of a vessel while intoxicated
192.5(b)	Vehicular manslaughter in the operation of a vessel while intoxicated
192.5(c)	Vehicular manslaughter in the operation of a vessel
203	Mayhem
205	Aggravated mayhem
206	Torture
207(a)	Kidnapping
207(b)	Kidnap - 14 to com l&l
207(c)	Kidnapping by false pretense
207(d)	Kidnapping from outside the state
208(b)	Kidnap child under 14 yrs
209(a)	Kidnapping for ransom
209(b)(1)	Kidnap: commit rob/rape/etc.
209.5(a)	Kidnap during carjacking
210.5	False imprisonment of a hostage
667.85	Kidnap to deprive parent
211	Robbery: first or second degree
212	Fear defined for robbery
212.5	Robbery; degrees
214	Train robbery
215	Carjacking
217.1(a)	Assault on a public official
217.1(b)	Attempted murder of a public official
218	Train wrecking; attempt; punishment
218.1	Obstructing railroad track; punishment

PENAL CODE	DESCRIPTION
219	Train derailing or wrecking; punishment
219.1	Throwing missile at common carrier with bodily harm
219.2	Throwing hard substance or shooting missile at train or other conveyance
220	Assault with intent to commit mayhem, rape, sodomy, oral copulation, or any violation of sections 264.1, 288, or 289
220(a)(1)	Assault with intent to commit a felony
220(a)(2)	Assault with intent to commit a felony-victim under 18
220(b)	Assault to commit a felony during the commission of a first degree burglary
222	Administering to another any chloroform, ether, laudanum, or any controlled substance, anesthetic, or intoxicating agent
236	False imprisonment
236.1	Human trafficking; provisions regarding minors; consideration of total circumstances
237(a)	False imprisonment
240	Assault
241	Assault
241.1	Assault on custodial officer
241.2	Assault on school or park property
241.3	Assault against person on public transportation, both on property of and within motor vehicle of provider
241.4	Assault on peace officer of a school district
241.5	Assault on a highway worker
241.6	Battery on school employee
241.7	Assault against jurors
241.8(a)	Battery against member of U.S. armed forces
242	Battery
243	Battery
243.1	Battery on custodial officer
243.2(a)(1)	Battery on pers on school/park/grnds
243.25	Battery on an elder or dependent adult
243.3	Battery on transportation personnel/passenger
243.35	Battery on public transportation provider
243.4	Sexual battery
243.5(a)(1)	Assault or battery on school prop
243.6	Battery on school employee
243.65(a)	Battery against a highway worker
243.7	Battery against jurors
243.8(a)	Battery against a sports official
243.9(a)	Aggravated battery by gassing on peace officer or local detention facility employee
244	Aslt w/caustic chem/etc.
244.5(b)	Assault with stun gun/taser
244.5(c)	Assault with stun gun or taser on peace officer or firefighter
245(a)(1)	Force/adw-not firearm: gbi
245(a)(2)	Aslt w/ firearm on person
245(a)(3)	Aslt w/machine gun on person
245(a)(4)	Force/adw not firearm: gbi
245(b)	Assault w/semiauto rifle
245(c)	Adw not f/arm: po/fire: gbi
245(d)(1)	Assault with a firearm upon a peace officer or firefighter

PENAL CODE	DESCRIPTION
245(d)(2)	Assault on peace officer/firefighter with semiautomatic firearm
245(d)(3)	Machine gun/assault weapon on a peace officer/firefighter
245.2	Assault (adw/gbi) upon transportation personnel, mass transit personnel
245.3	Assault (adw/gbi) upon a custodial officer
245.5(a)	Adw/gbi schl emp: no f/arm
245.5(b)	Assault with firearm on a school employee
245.5(c)	Adw/stun gun or taser: school employee
245.6	Hazing resulting in death/serious bodily injury
246	Shoot: inhab dwell/veh/etc.
246.3(a)	Firearm disch w/neg
246.3(b)	BB device disch w/neg
261(a)	Rape
261.5(a)	Sex intercourse w/mnr - 18
261.5(b)	Sex w/minor: + or - 3 yrs
261.5(c)	Sex w/minor:3+ yrs younger
261.5(d)	Sex w/minor: perp 21 + vic-16
262(a)(1)	Rape spouse by force/etc.
262(a)(2)	Rape spouse und c/sub/etc.
262(a)(3)	Rape: spouse uncon of act
262(a)(4)	Rape: spouse - threat to kidnap, inflict extreme pain, serious bodily injury
262(a)(5)	Rape: spouse - threat to incarcerate, arrest, deport
262(a)(6)	Rape of spouse by threat to arrest or deport
264.1	Rape/etc.: cncrt force/viol
266a	Taking a person for prostitution
266b	Abduction to live in illicit relation; using force
266c	Unlawful sexual intercourse, sexual penetration, oral copulation, or sodomy; consent procured by false or fraudulent representation with intent to create fear
266h(b)	Pimping a minor
266i(b)	Pandering a minor
266j	Procurement of child under age 16 for lewd and lascivious acts
267	Abduction; person under 18 for purpose of prostitution
269(a)	Agg sex aslt: mnr: frce/etc.
273.4	Female genital mutilation
273.5(a)	Injuring a spouse, cohabitant, fiancé, boyfriend, girlfriend, or child's parent
273.5(f)	Inf crpl inj: sps/etc. w/pr
273.6(b)	Viol crt ord to prev domes viol – results in physical injury
273.6(d)	Domestic violence w/prior – act of violence or a credible threat of violence
273a(a)	Willful cruelty to child/poss inj/death
273a(b)	Willful cruelty to child
273ab(a)	Assault of child under 8 by force likely to produce GBI resulting in death
273ab(b)	Assault of child under 8 by force likely to produce GBI resulting in brain injury, paralysis
273d(a)	Inflict injury upon child
278	Child stealing
285	Incest
286(b)	Sodomy: person under 18
286(c)	Sodomy: person under 14
286(d)	Sodomy in concert w/force
286(f)	Sodomy: vict uncons of act

PENAL CODE	DESCRIPTION
286(g)	Sodomy: vict incapbl:consent
286(h)	Sodomy: vic/def in mntl inst
286(i)	Sodomy: no ok: vict drugged
286(j)	Sodomy by impersonation
286(k)	Sodomy under color of authority
288(a)	Lewd or lasciv acts/w/child under 14yrs
288(b)	Lewd/lasc acts w/child under 14 or dependent person
288(c)	Lewd/lasc act w/chld 14/15: def 10yr+ or dependent person
288.2(a)	Harmful mtr sent w/int of seducing minor
288.3	Contact with intent to commit sex act
288.4	Arranging a meeting with minor for lewd purposes
288.5(a)	Continuous sexual abuse of child
288.7(a)	Sex/sodomy with a child under 10
288.7(b)	Oral copulation/sexual penetration with a child under 10
287(b)	Oral copulation w/pers under 18yrs
287(c)	Oral copul w/person under 14/by force
287(d)	Oral cop in concert: vic incap of con
287(f)	Oral cop: vic uncon/asleep
287(g)	Oral copulation of an incompetent person
287(h)	Oral cop: vic/def in mntl inst
287(i)	Oral copulation by anesthesia or controlled substance
287(j)	Oral copulation by impersonation
287(k)	Oral copulation under color of authority
288a(b)	Oral copulation w/person under 18yrs
288a(c)	Oral copul w/person under 14/by force
288a(d)	Oral cop in concert: vic incap of con
288a(f)	Oral cop: vic uncon/asleep
288a(g)	Oral copulation of an incompetent person
288a(h)	Oral cop: vic/def in mntl inst
288a(i)	Oral copulation by anesthesia or controlled substance
288a(j)	Oral copulation by impersonation
288a(k)	Oral copulation under color of authority
289	Sexual pen with force/etc.
289.6(a)(3)	Sex: emp/etc. cnf/detention fac
311.4(a)	Using Minors for Sex Acts
311.4(b)	Using Minors for Commercial Sex Acts
311.4(c)	Using Minors for Sex Acts
347(a)	Poisoning, willful poison/etc. food/etc.
368(b)	Cause harm/death elder dep adult
368(c)	Elder/dependent adult cruelty
368(f)	False imprison: elder/dep adult violence
404(a)	Rioting
417(a)	Exhibit firearm or deadly weapon other than gun. Drawing, exhibiting, or using firearm or deadly weapon; self-defense; peace officers.
417(b)	Exhibit firearm. Drawing, exhibiting, or using a firearm.
417(c)	Exhibit firearm in presence of p.o. Drawing, exhibiting, or using firearm or deadly weapon; self-defense; peace officers.
417.3	Exhibit firearm pres beh occupt
417.8	Exhibit firearm/etc.: resist arrest
422.6(a)	Violate civil rights by force or threat
451(a)	Arson causing great bodily injury

PENAL CODE	DESCRIPTION
451(b)	Arson: inhabited structure/property
451.1	Arson with added circumstances
451.5(a)	Aggravated arson
452(a)	Causing fire that causes gbi
452(b)	Causing fire of inhabited struc/prop
455	Arson attempts and acts preliminary or in furtherance
646.9(a)	Stalking
646.9(b)	Stalking/temp restraining order
647.6(a)(1)	Annoy/molest child under 18yrs
647.6(b)	Annoy/molest child/ill entry of bldg
647.6(c)	Annoy/etc. child - 18 w/prior
667.61(d)(2)	Felony sex offenses; victim kidnapped increasing risk of harm
667.61(d)(3)	Felony sex offenses; victim tortured
667.61(e)(1)	Felony sex offense; victim kidnapped
667.61(e)(2)	Felony sex offenses during commission of burglary
667.61(e)(4)	Felony sex offenses against more than one victim
667.61(e)(5)	Felony sex offenses - tying or binding of victim or another person
667.8	Kidnap to commit sex offense
667.85	Kidnap child under 14 yrs
674	Sex offense by daycare provider
836.6(c)	Escape from custody by force or violence
4500	Assault by a life prisoner
4501	Assault by a state prisoner
4501.1(a)	Aggravated battery
4501.5	Battery on nonconfined person by prisoner
4503	Holding of hostages; offense
4530(a)	Escape from custody by force and violence
4532(a)(2)	Escape from alternative custody by force or violence by person booked on misdemeanor
4532(b)(2)	Escape from alternative custody by force or violence by person booked on felony
11413(a)	Terrorism by explosion
11413(b)	Terrorism by explosion (specified places)
11418(b)	Weapons of mass destruction: use and damage to life
11418(c)	Weapons of mass destruction: use and damage to public natural resources
11418(d)	Weapons of mass destruction: creation of new pathogens
18740	Use of destructive device and explosive to injure/destroy
18745	Explosion with intent to murder
18750	Explosion of destructive device causing bodily injury
18755	Explosion causing death, mayhem, GBI
26100(c)	Discharge of firearm at another person from motor vehicle
18540(a)	Use of firearm to intimidate a voter
664/187(a)	Attempted murder
664/211	Attempted robbery
Veh. Code, § 2800.3(a)	SBI caused by flight from peace officer
Veh. Code, § 2800.3(b)	Death caused by flight from peace officer
All attempts (Pen. Code, § 664), conspiracy (Pen. Code, § 182), solicitation (Pen. Code, § 653f), and accessory (Pen. Code, § 31) only if before the act of any of the offenses identified here also meet the definition of a violent offense for purposes of administering the PSA.	

APPENDIX C

TABLE C1. PRETRIAL PILOT PROGRAM DATE RANGE BY COUNTY

County	Earliest Assessment Date	Latest Assessment Date
Alameda	2020-05-12	2022-02-03
Calaveras	2019-10-10	2021-03-23
Kings	2020-07-24	2021-12-22
Los Angeles	2020-03-23	2021-12-31
Modoc	2020-07-27	2021-06-03
Napa	2020-01-15	2021-12-16
Nevada	2020-06-02	2021-07-16
Sacramento	2019-10-21	2022-07-13
San Joaquin	2018-01-01	2022-01-05
San Mateo	2020-01-30	2022-01-01
Sonoma	2020-07-01	2022-01-10
Tulare	2018-07-03	2022-04-27
Tuolumne	2021-07-01	2022-01-01
Ventura	2019-10-01	2022-02-23
Yuba	2020-01-02	2022-01-31

APPENDIX D

PSA AUC Race/Ethnicity Comparison

```
 DeLong's test for two ROC curves

data:  rocW and rocB
D = 1.179, df = 18319, p-value = 0.2384
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6671416   0.6562089


 DeLong's test for two ROC curves

data:  rocW and rocH
D = -3.685, df = 18065, p-value = 0.0002293
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6671416   0.6964127


 DeLong's test for two ROC curves

data:  rocB and rocH
D = -4.9177, df = 18785, p-value = 8.829e-07
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6562089   0.6964127

 DeLong's test for two ROC curves

data:  rocW and rocB
D = 4.8702, df = 18296, p-value = 1.124e-06
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7021626   0.6643501


 DeLong's test for two ROC curves

data:  rocW and rocH
D = -5.6352, df = 16744, p-value = 1.777e-08
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7021626   0.7392599


 DeLong's test for two ROC curves

data:  rocB and rocH
D = -11.311, df = 17994, p-value < 2.2e-16
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6643501   0.7392599
```

```
 DeLong's test for two ROC curves

data:  rocW and rocB
D = 1.5888, df = 18307, p-value = 0.1121
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6624258   0.6470699


 DeLong's test for two ROC curves

data:  rocW and rocH
D = -5.0086, df = 16885, p-value = 5.537e-07
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6624258   0.7023190


 DeLong's test for two ROC curves

data:  rocB and rocH
D = -6.608, df = 17243, p-value = 4.008e-11
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6470699   0.7023190


 DeLong's test for two ROC curves

data:  rocW and rocB
D = 1.5439, df = 18311, p-value = 0.1226
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6579199   0.6410380


 DeLong's test for two ROC curves

data:  rocW and rocH
D = -4.523, df = 16197, p-value = 6.14e-06
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6579199   0.6981747


 DeLong's test for two ROC curves

data:  rocB and rocH
D = -6.1256, df = 16641, p-value = 9.236e-10
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6410380   0.6981747


 DeLong's test for two ROC curves

data:  rocW and rocB
D = -0.055549, df = 18011, p-value = 0.9557
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.5784651   0.5790339
```

```
 DeLong's test for two ROC curves

data:  rocW and rocH
D = 0.36365, df = 16948, p-value = 0.7161
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.5784651   0.5751451


 DeLong's test for two ROC curves

data:  rocB and rocH
D = 0.45131, df = 19760, p-value = 0.6518
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.5790339   0.5751451
```

PSA AUC Gender Comparison

```
 DeLong's test for two ROC curves

data:  rocW_FTA and rocM_FTA
D = 1.0898, df = 9141.2, p-value = 0.2758
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
   0.693214    0.683843


 DeLong's test for two ROC curves

data:  rocW_newarrest and rocM_newarrest
D = -0.074471, df = 8948.8, p-value = 0.9406
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7101721   0.7107066


 DeLong's test for two ROC curves

data:  rocW_newfiling and rocM_newfiling
D = -0.44239, df = 8548.2, p-value = 0.6582
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6742954   0.6784097
```

```
 DeLong's test for two ROC curves

data:  rocW_newconviction and rocM_newconviction
D = 0.22765, df = 8427.1, p-value = 0.8199
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6748149   0.6724270


 DeLong's test for two ROC curves

data:  rocW_newviolent and rocM_newviolent
D = 0.3086, df = 9025.3, p-value = 0.7576
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.5824270   0.5794453
```

VPRAI AUC Race/Ethnicity Comparison

```
 DeLong's test for two ROC curves

data:  rocW and rocB
D = 2.4905, df = 2551.1, p-value = 0.01282
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6967694   0.6470130


 DeLong's test for two ROC curves

data:  rocW and rocH
D = 0.45646, df = 3660.4, p-value = 0.6481
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6967694   0.6889505


 DeLong's test for two ROC curves

data:  rocB and rocH
D = -2.1394, df = 2641.9, p-value = 0.03249
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6470130   0.6889505
```

```
 DeLong's test for two ROC curves

data:  rocW and rocB
D = 2.4677, df = 2642.5, p-value = 0.01366
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7483248   0.7028127


 DeLong's test for two ROC curves

data:  rocW and rocH
D = -0.018816, df = 3430, p-value = 0.985
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7483248   0.7486181

 DeLong's test for two ROC curves

data:  rocB and rocH
D = -2.6477, df = 2523.1, p-value = 0.008155
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7028127   0.7486181

 DeLong's test for two ROC curves

data:  rocW and rocB
D = 0.57193, df = 2624, p-value = 0.5674
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6812106   0.6699262


 DeLong's test for two ROC curves

data:  rocW and rocH
D = -2.447, df = 3511.2, p-value = 0.01445
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6812106   0.7223674

 DeLong's test for two ROC curves

data:  rocB and rocH
D = -2.7932, df = 2577.5, p-value = 0.005257
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6699262   0.7223674
```

```
 DeLong's test for two ROC curves

data:  rocW and rocB
D = 0.28112, df = 2576.8, p-value = 0.7786
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6695253   0.6635067


 DeLong's test for two ROC curves

data:  rocW and rocH
D = -2.4451, df = 3506.7, p-value = 0.01453
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6695253   0.7132254


 DeLong's test for two ROC curves

data:  rocB and rocH
D = -2.4374, df = 2501, p-value = 0.01486
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6635067   0.7132254

 DeLong's test for two ROC curves

data:  rocW and rocB
D = -0.83233, df = 2802, p-value = 0.4053
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6082175   0.6275826
```

```
 DeLong's test for two ROC curves

data:  rocW and rocH
D = -2.9265, df = 3307, p-value = 0.003451
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6082175   0.6693782


 DeLong's test for two ROC curves

data:  rocB and rocH
D = -1.9825, df = 2738, p-value = 0.04752
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6275826   0.6693782

 DeLong's test for two ROC curves

data:  rocW_anyfailure and rocB_anyfailure
D = 3.3701, df = 2621.9, p-value = 0.0007624
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7596811   0.6966753

 DeLong's test for two ROC curves

data:  rocW_anyfailure and rocH_anyfailure
D = 1.4538, df = 3447, p-value = 0.1461
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7596811   0.7368354


 DeLong's test for two ROC curves

data:  rocB_anyfailure and rocH_anyfailure
D = -2.2812, df = 2507.8, p-value = 0.02262
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6966753   0.7368354
```

VPRAI AUC Gender Comparison

```
 DeLong's test for two ROC curves

data:  rocW_FTA and rocM_FTA
D = -0.35531, df = 1705.9, p-value = 0.7224
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6873398   0.6937618

 DeLong's test for two ROC curves

data:  rocW_newarrest and rocM_newarrest
D = -1.7491, df = 1673, p-value = 0.08045
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7206748   0.7494749

 DeLong's test for two ROC curves

data:  rocW_newfiling and rocM_newfiling
D = -1.2977, df = 1592.3, p-value = 0.1946
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6829305   0.7076166


 DeLong's test for two ROC curves

data:  rocW_newconviction and rocM_newconviction
D = -1.4972, df = 1512.6, p-value = 0.1346
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6653929   0.6981285


 DeLong's test for two ROC curves

data:  rocW_newviolent and rocM_newviolent
D = -0.77399, df = 1501, p-value = 0.4391
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6244547   0.6435426


 DeLong's test for two ROC curves

data:  rocW_anyfailure and rocM_anyfailure
D = -1.1546, df = 1708.4, p-value = 0.2484
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7251835   0.7440326
```

VPRAI-R AUC Race/Ethnicity Comparison

```
 DeLong's test for two ROC curves

data:  rocW_FTA and rocB_FTA
D = 1.6599, df = 1874.1, p-value = 0.0971
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6653762   0.6253172


 DeLong's test for two ROC curves

data:  rocW_FTA and rocH_FTA
D = -1.6317, df = 1794.1, p-value = 0.1029
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6653762   0.7034199


 DeLong's test for two ROC curves

data:  rocB_FTA and rocH_FTA
D = -3.6586, df = 2675.5, p-value = 0.0002585
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6253172   0.7034199
```

```
 DeLong's test for two ROC curves

data:  rocW_newarrest and rocB_newarrest
D = 2.399, df = 1849.9, p-value = 0.01654
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7187533   0.6635681


 DeLong's test for two ROC curves

data:  rocW_newarrest and rocH_newarrest
D = -0.89307, df = 1694.4, p-value = 0.3719
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7187533   0.7382471


 DeLong's test for two ROC curves

data:  rocB_newarrest and rocH_newarrest
D = -3.798, df = 2634.8, p-value = 0.0001492
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6635681   0.7382471


 DeLong's test for two ROC curves

data:  rocW_newfiling and rocB_newfiling
D = 2.2853, df = 1913.5, p-value = 0.02241
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6891515   0.6312669


 DeLong's test for two ROC curves

data:  rocW_newfiling and rocH_newfiling
D = 0.87724, df = 1910.3, p-value = 0.3805
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6891515   0.6672689


 DeLong's test for two ROC curves

data:  rocW_newconviction and rocB_newconviction
D = 2.1488, df = 2013.4, p-value = 0.03177
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6795214   0.6147869
```

```
 DeLong's test for two ROC curves

data:  rocW_newconviction and rocH_newconviction
D = 1.0979, df = 2027.6, p-value = 0.2724
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6795214   0.6470562


 DeLong's test for two ROC curves

data:  rocB_newconviction and rocH_newconviction
D = -1.1075, df = 2701.1, p-value = 0.2682
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6147869   0.6470562


 DeLong's test for two ROC curves

data:  rocW_newviolent and rocB_newviolent
D = 0.63909, df = 1678, p-value = 0.5228
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.5784692   0.5578102


 DeLong's test for two ROC curves

data:  rocW_newviolent and rocH_newviolent
D = -1.0112, df = 1746.6, p-value = 0.312
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.5784692   0.6115525

 DeLong's test for two ROC curves

data:  rocB_newviolent and rocH_newviolent
D = -1.9473, df = 2718.8, p-value = 0.0516
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.5578102   0.6115525


 DeLong's test for two ROC curves

data:  rocW_anyfailure and rocB_anyfailure
D = 3.0018, df = 1860.1, p-value = 0.002719
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7256606   0.6555694
```

```
 DeLong's test for two ROC curves

data:  rocW_anyfailure and rocH_anyfailure
D = -1.0478, df = 1657, p-value = 0.2949
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7256606   0.7485816


 DeLong's test for two ROC curves

data:  rocB_anyfailure and rocH_anyfailure
D = -4.7084, df = 2600.5, p-value = 2.628e-06
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6555694   0.7485816
```

VPRAI-R AUC Gender Comparison

```
 DeLong's test for two ROC curves

data:  rocW_FTA and rocM_FTA
D = -1.2178, df = 856.11, p-value = 0.2236
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6461734   0.6759365


 DeLong's test for two ROC curves

data:  rocW_newarrest and rocM_newarrest
D = -2.5388, df = 843.39, p-value = 0.0113
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6678574   0.7259254


 DeLong's test for two ROC curves

data:  rocW_newfiling and rocM_newfiling
D = -1.0744, df = 807.58, p-value = 0.283
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6387910   0.6696055
```

```
 DeLong's test for two ROC curves

data:  rocW_newconviction and rocM_newconviction
D = -1.1592, df = 790.99, p-value = 0.2467
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6168450   0.6585583


 DeLong's test for two ROC curves

data:  rocW_newviolent and rocM_newviolent
D = 0.15986, df = 850.53, p-value = 0.873
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.5804364   0.5752361


 DeLong's test for two ROC curves

data:  rocW_anyfailure and rocM_anyfailure
D = -2.4917, df = 844.38, p-value = 0.0129
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6718484   0.7290131
```

ORAS AUC Race/Ethnicity Comparison

```
 DeLong's test for two ROC curves

data:  rocW_FTA and rocB_FTA
D = 0.065327, df = 275.62, p-value = 0.948
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6596060   0.6567832


 DeLong's test for two ROC curves

data:  rocW_FTA and rocH_FTA
D = 0.027929, df = 2624.9, p-value = 0.9777
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6596060   0.6590036
```

```
 DeLong's test for two ROC curves

data:  rocB_FTA and rocH_FTA
D = -0.052407, df = 255.55, p-value = 0.9582
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6567832   0.6590036


 DeLong's test for two ROC curves

data:  rocW_newarrest and rocB_newarrest
D = 1.5157, df = 262.05, p-value = 0.1308
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7236282   0.6648417


 DeLong's test for two ROC curves

data:  rocW_newarrest and rocH_newarrest
D = 0.23823, df = 2623.1, p-value = 0.8117
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7236282   0.7193955


 DeLong's test for two ROC curves

data:  rocB_newarrest and rocH_newarrest
D = -1.4301, df = 245.7, p-value = 0.1539
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6648417   0.7193955

 DeLong's test for two ROC curves

data:  rocW_newfiling and rocB_newfiling
D = -0.068749, df = 256.33, p-value = 0.9452
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6999268   0.7034309
```

```
 DeLong's test for two ROC curves

data:  rocW_newfiling and rocH_newfiling
D = 0.58554, df = 2815, p-value = 0.5582
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6999268   0.6860022


 DeLong's test for two ROC curves

data:  rocB_newfiling and rocH_newfiling
D = 0.34278, df = 254.1, p-value = 0.7321
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7034309   0.6860022


 DeLong's test for two ROC curves

data:  rocW_newconviction and rocB_newconviction
D = -0.22516, df = 245.95, p-value = 0.822
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7059287   0.7197917


 DeLong's test for two ROC curves

data:  rocW_newconviction and rocH_newconviction
D = 0.42288, df = 2765.8, p-value = 0.6724
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7059287   0.6950706


 DeLong's test for two ROC curves

data:  rocB_newconviction and rocH_newconviction
D = 0.40367, df = 240.92, p-value = 0.6868
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7197917   0.6950706
```

```
 DeLong's test for two ROC curves

data:  rocW_newviolent and rocB_newviolent
D = 0.089947, df = 247.63, p-value = 0.9284
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:

 DeLong's test for two ROC curves

data:  rocW_newviolent and rocH_newviolent
D = 0.16728, df = 2633.6, p-value = 0.8672
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6327517   0.6275507


 DeLong's test for two ROC curves

data:  rocB_newviolent and rocH_newviolent
D = -0.022054, df = 235.77, p-value = 0.9824
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6258890   0.6275507
```

ORAS AUC Gender Comparison

```
 DeLong's test for two ROC curves

data:  rocW_FTA and rocM_FTA
D = -0.9763, df = 1147.3, p-value = 0.3291
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6408353   0.6651606


 DeLong's test for two ROC curves

data:  rocW_newarrest and rocM_newarrest
D = -0.46466, df = 1148.2, p-value = 0.6423
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.7163807   0.7258438


 DeLong's test for two ROC curves

data:  rocW_newfiling and rocM_newfiling
D = -0.14203, df = 1050.4, p-value = 0.8871
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6958142   0.7000722
```

```
 DeLong's test for two ROC curves

data:  rocW_newconviction and rocM_newconviction
D = -0.27594, df = 981.99, p-value = 0.7827
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6971687   0.7069161


 DeLong's test for two ROC curves

data:  rocW_newviolent and rocM_newviolent
D = 0.20765, df = 1211.1, p-value = 0.8355
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.6397206   0.6324394
```