Contra Costa County Probation Department

SB 36 Public Safety Assessment Validation

RDA CONSULTING

40 YEARS of Systems Change

Contra Costa County Probation

Public Safety Assessment (PSA) Validity Study

This report was developed by RDA Consulting under contract with the Contra Costa County Probation Department.

RDA Consulting, March 2025

Contents

Introduction

Applying the principles of effective intervention, or the risk, need, responsivity (RNR) model in corrections, reduces the likelihood that an individual will reoffend. The RNR model is best executed through a validated risk and needs assessment (RNA) tool. In 2018, Arnold Ventures released the Public Safety Assessment (PSA) as a free pretrial assessment tool that can be implemented without an interview and predicts three distinct outcomes: failure to appear (FTA), new criminal activity (NCA), and new violent criminal activity (NVCA). Since initial implementation, the PSA tool has been adopted in more than 22 counties throughout California, including Alameda, Los Angeles, Sacramento, Santa Clara, and Riverside.

RNAs such as the PSA cannot be valid if they are not reliable at the local level. However, when applied to jurisdictions outside the development sample, instruments become vulnerable to performance or prediction shrinkage (i.e., declines in predictive accuracy or validity) as individuals' characteristics, laws, agency policies, and local supervision conditions inevitably vary from jurisdiction to jurisdiction. Additionally, RNAs must be fair to all assessed subpopulations, meaning instruments should evaluate racial and ethnic minorities and socioeconomically disadvantaged individuals equitably to ensure all groups receive supervision and treatment that corresponds with their actual levels of risk.

In compliance with Senate Bill 36 (SB 36), the Contra Costa County Probation Department (the Probation Department) contracted with RDA in May 2024 to complete a validation of its pretrial risk assessment. This study is the first external evaluation to validate local use of the PSA since its adoption in the County. This validation aims to determine if the tool accurately predicts risk (e.g., failure to appear in Court and new criminal activity) and to understand and reduce any detectable biases based on gender, race, or ethnicity in pretrial release decision-making. It is important to regularly re-evaluate the PSA's local predictive validity following this initial validation study to adhere to risk assessment validation best practices and SB 36's requirements.

The PSA in Contra Costa County

After previously using the Virginia Pretrial Risk Assessment Instrument, Contra Costa County adopted the PSA in 2020 to help its Adult Services and Court staff make more informed decisions at critical points in the criminal justice system process.

Probation Officers in the Department's Pretrial Services Unit administer the PSA for every eligible defendant held in custody to determine the suitability of release before sentencing and the general level of supervision during their pretrial release.[1] The PSA contains nine risk factors displayed in **Table i** to predict the likelihood that a defendant released pretrial will (1) fail to appear in Court for their sentencing (i.e., FTA), (2) commit a new criminal offense (i.e., NCA), and (3) commit a new violent criminal offense (i.e., NVCA). In summary, the tool evaluates known risk factors such as the defendant's age, criminal conviction history, history of appearing in Court, and their current (i.e., instant) offense to predict a defendant's likelihood of FTA, NCA, and NVCA.

[1] Defendants held in custody are not eligible for pretrial release if they meet any of the following conditions: (1) have to register under Penal Code Section 290, (2) currently supervised on Post Release Community Supervision (PRCS), (3) currently supervised on Mandatory Supervision, (4) currently has an out-of-county felony warrant, (5) in custody for a current offense that is serious or violent, under Penal Code Section (§) 667.5, (6) in custody for a current offense that is serious or violent, under Penal Code § 1192.7.

Table i. PSA Risk Factors Corresponding with Each Predicted Outcome

Risk Factor	FTA	NCA	NVCA
1. Age (22 or younger) at arrest		X	
2. Current violent offense			X
2a. Current violent offense & 20 years old or younger			X
3. Pending charge at time of offense	X	X	X
4. Prior misdemeanor conviction		X	
5. Prior felony conviction		X	
5a. Prior conviction (misdemeanor or felony)	X		X
6. Prior violent conviction		X	X
7. Prior failure to appear in past two years	X	X	
8. Prior failure to appear older than two years	X		
9. Prior sentence to incarceration		X	

Officers score risk factors using a records search without a face-to-face interview. After reviewing available justice data (e.g., California Law Enforcement Telecommunications System, court documents, and Probation case management systems), risk factors for each predicted outcome are added together by Probation's assessment system with applicable weights to generate scaled FTA and NCA scores along with a "yes/no" NVCA flag. These FTA and NCA scaled scores take values of one to six, with one being least likely to have the predicted outcome.

Officers provide the Court with PSA assessment results in a written pretrial report, including a defendant's scaled scores, individual risk factors, and release recommendations. Reports include additional information regarding the defendant, such as booking charges, the circumstances of the instant offense, and the defendant's release level derived from **Figure i**. Probation utilizes the release level from this decision-making framework in conjunction with a recommendation matrix and an NVCA flag to produce Probation's release recommendation and supervision term recommendations (e.g., frequency of check-ins).

Ultimate release decision authority rests with the Judicial Officer of the Court, who can deny pretrial release, release defendants on their own recognizance, or release defendants under the

Table ii. Contra Costa County Decision Making Framework (2020-2023)

	NCA 1	NCA 2	NCA 3	NCA 4	NCA 5	NCA 6
FTA 1	1	1				
FTA 2	1	1	1	2	3	
FTA 3		2	2	2	3	4
FTA 4		2	3	3	4	4
FTA 5		2	3	4	4	4
FTA 6				4	4	4

supervision of Probation's Pretrial Services Unit. Individuals released on their own recognizance are required to appear in Court but typically do not have any Probation reporting requirements or other terms for their release. When a defendant is not deemed appropriate to be released on their own recognizance by the Judicial Officer, defendants are placed on supervised release with terms and conditions (e.g., office visits and GPS monitoring) to ensure they appear in Court.

Current Study

The County contracted with RDA in May 2024 to complete a validation of its pretrial risk assessment tool to share with the public in compliance with the requirements of SB 36. The goals of the project are to determine if the tool is accurately predicting defendants' risk to recidivate (i.e., failure to appear in Court, new criminal activity, and new violent criminal activity) and understand and reduce any detectable biases based on gender and race/ethnicity in pretrial release decision-making. Specifically, the following objectives guided the study design:

Objective 1. Learn about local assessment policies and practices through survey data collection with probation officers.

Objective 2. Investigate predictive validity or the extent to which the PSA correctly measures what it is supposed to measure — in this case, the likelihood of a pretrial failure.

Subsequent sections of this report define the methods used to meet these objectives, describe the study findings, and conclude with a final summative discussion of the study's findings and instrument and agency improvement recommendations.

I. Methodology

To evaluate the effectiveness of the PSA, RDA followed the data collection, sampling, and analysis methodologies outlined in this section. This study employed a mixed-methods approach to data collection and analysis, integrating staff perspectives with quantitative rigor to paint a complete picture regarding assessment practices and implementation while contextualizing predictive validity results to provide more precise recommendations.

Local Assessment Practices

To understand local assessment practices and how the PSA is used in the County, RDA developed and administered officer surveys to explore the implementation of the tools, including data sources officers consult as part of the risk assessment, the usefulness of the PSA, and officer confidence in the PSA results. Additionally, RDA spoke with Probation administrators to understand department policies and day-to-day practices. Qualitative data findings from the survey and quantitative findings from officer surveys help RDA and the County identify facilitators and barriers to PSA assessment implementation and administration.

Recruitment. County Probation administrators identified six officers in the Pretrial Services Unit who had experience administering the PSA. Probation administrators sent an introductory e-mail to the selected survey participants informing them about the evaluation and when to anticipate an e-mail from RDA with a unique survey link. RDA initiated outreach on July 15, 2024, and all officers completed the survey within two days.

Questions. RDA developed a survey tool with input from Probation administrators to understand how the PSA is conducted and identify the tool's strengths and challenges. The complete survey tool can be reviewed in **Appendix A**. Specific questions included whether assessors refer to scoring guides when utilizing the PSA, whether assessors use the release conditions matrix to determine a defendant's recommended release level, and whether there is a quality assurance process for assessment scoring. RDA used Alchmer, a secure, web-based survey platform, to administer and collect survey responses from officers.

Analysis. Qualitative data collected from open-ended officer responses were analyzed using directed content analysis strategies to identify and code themes identified by common words and similar explicit or inferred meanings within participants' responses. Additionally, responses on themes were grouped to allow for some basic quantification (e.g., this many respondents referred to a particular "strength" or "limitation" when administering the PSA or calculating assessment results). Quantitative data from the survey were analyzed to produce descriptive statistics such as frequencies and averages.

PSA Predictive Validity

To evaluate the predictive accuracy of the PSA, RDA collected administrative data from the Probation Department related to defendants who received assessments, their assessment scores, and outcomes. The predictive validity sample spanned four years and included defendants with PSA assessments completed between 2020 and 2023. Analyses related to the evaluation of the psychometric properties of the instruments and normative comparisons helped RDA determine whether PSA items are theoretically relevant and connected with the measurement of criminogenic risk. The sample consisted of defendants who were held in custody and released before sentencing.

Sampling Strategy. The Probation Department provided data with defendants' names redacted for this validation study. Specifically, Probation provided an Excel workbook with three tabs of

data, including PSA results, pretrial outcomes, Probation's release recommendations, and judicial officer release decisions. The evaluation team joined these datasets using "Person ID" and "Case ID" fields that identified individuals and distinct cases across all tabs provided.

The evaluation team selected each defendant's earliest assessment for analyses to align with statistical modeling assumptions.[2] The validation sample was next limited from 2,068 defendants with a PSA assessment to the 1,114 defendants who were released pretrial, including defendants on supervised release and defendants released on their own recognizance.[3] The sample was finally limited to the 990 individuals who had completed their pretrial released term.

Ultimately, the Department completed PSA assessments for 990 distinct released defendants between 2020 and 2023 based on the exclusion criteria described in the preceding paragraphs. Displayed in **Table 1** are the characteristics of these 990 released defendants included in the PSA predictive validity study sample. Sample sizes must be large enough to support cross-tabulations of discrete assessment groups and subgroups of interest within each assessment sample. If too few cases are collected, the number of observations in analytical subgroups will fall below levels supporting meaningful analysis. In short, larger samples will support more reliable and informative findings.[4]

Predictor & Outcome Variables. Predictor variables for this study include the results of a defendant's earliest matched assessment, including, for example, factor scores, scaled scores, and an NVCA flag. Dichotomous recidivism measures, including (1) failure to appear in Court (FTA) as required, (2) new criminal activity (NCA), and (3) new violent criminal activity (NVCA), were created to determine the predictive accuracy and fairness the PSA.[5]

---

[2] The research team also retained each defendant's most serious charge (i.e., felony and enhancements).

[3] Regardless of their supervision status, all defendants released pretrial could have "failed" by not appearing in Court, engaging in new criminal activity, or a new violent criminal activity. This study aims to assess the PSA's ability to correctly predict any pretrial failure, not the effectiveness of supervision in preventing a failure. Therefore, all released defendants meeting the study exclusion criteria are included in the primary analysis shared in the body of this report. However, to examine if limiting the sample to only those defendants on a supervised caseload changed the validation findings, the evaluation team conducted another set of analyses included in **Appendix J**. AUC values improved slightly for each outcome when limiting to the supervised population, but predictive accuracy remained fair for FTA (0.60), slight for NCA (0.54), and weak for NVCA (0.49).

[4] Mamalian, C. A. (2011). State of the science of pretrial risk assessment. Jointly published by the Pretrial Justice Institute and the Bureau of Justice Assistance. Washington D.C. Retrieved from https://bja.ojp.gov/sites/g/files/xyckuh186/files/Publications/PJI_PretrialRiskAssessment.pdf

[5] Defendants released pretrial may have their pretrial release revoked due to a failure to comply with their supervision terms; however, the PSA does not explicitly predict this type of failure. For this validation, the evaluation created an additional recidivism outcome measure, "FTA plus violation," that combines technical violations and other violations of supervision terms with FTAs. Predictive validity results for this outcome type are included in **Appendix I**. AUC values improved slightly when including violations, but predictive accuracy remained fair (0.60).

Table 1. Defendant Sample Characteristics (N = 990)

Characteristic	n	% or $\bar{x}$ (*SD*)
Gender		
Male	830	84.0%
Female	158	16.0%
Race		
White	265	26.8%
Hispanic	276	27.9%
Black	394	39.8%
Other	55	5.6%
Age		
18-26	261	26.4%
27-34	282	28.5%
35-44	266	26.9%
45-76	181	18.3%
Age	990	34.6 (11.3)
Instant Offense Included Violent Charge		
No	929	93.8%
Yes	61	6.2%
Contra Costa County Matrix Risk Level		
1	286	28.9%
2	297	30.0%
3	161	16.3%
4	246	24.8%
Matrix Risk Level	990	2.4 (1.1)
FTA Scaled Score	990	3.2 (1.5)
NCA Scaled Score	990	3.4 (1.3)
NVCA Flag		
0 (i.e., no flag)	960	97.0%
1 (i.e., NVCA flag)	30	3.0%

Statistical Analyses. All data underwent a preliminary round of data preparation in Excel before importing into Stata for statistical analyses. First, the team calculated univariate descriptive statistics for defendant demographic characteristics presented in **Table 1** and justice processing measures to describe the nature and distribution of the PSA assessment sample.

Second, the evaluation team normed the PSA instrument based on assessments completed in Contra Costa County. The norming process involved establishing the minimum and maximum total risk and item scores and calculating frequencies to examine the distribution of defendants falling into each scaled score category. The team also calculated average scaled scores for the complete sample of assessments and executed t-tests to investigate possible differences in assessed risk as a function of each subgroup of interest (e.g., race or ethnicity, gender).

To explore the internal reliability of the PSA, the evaluation team calculated Cronbach's alpha internal consistency coefficients. This measure determines the level of internal consistency between item scores to determine how closely related they are as a group. A coefficient score greater than or equal to .70 is typically considered acceptable. Additionally, the evaluation team calculated AUC-ROC (area under the receiver operating characteristic curve) and interpreted using the following minimum values: fair (AUC-ROC = .55), good (AUC-ROC = .64), and excellent (AUC-ROC = .71).[6] Of note, the AUC-ROC is not affected by samples' base rates of reoffending and is therefore considered a better measure of internal reliability than Pearson's r correlation coefficient.

The evaluation team examined the predictive validity of the PSA with basic frequencies and chi-square tests of significance to examine the relationship between assessed risk levels and outcomes (i.e., FTA, NCA, and NVCA). The team also utilized Pearson's r to examine the direction and relative strength of the relationship between assessment scores and recidivism. Pearson's r was interpreted using the following minimum values: fair (r = .10), good (r = .24), and excellent (r = .37).[7] The evaluation team used AUC-ROC and logistic regression to measure the PSA's ability to predict pretrial failure correctly. To further contextualize these results, the evaluation team replicated additional analyses utilized in the California Policy Lab's San Francisco County PSA validation, the evaluation team also examined correct risk classification on a binary scale.[8] , Specifically, the evaluation team assessed the predictive accuracy of individual risk factors and examined correct risk classification on a binary scale, across the complete validation sample and by different racial and ethnic identities.

Limitations

Several limitations could influence the results of this validation study. To ensure the validation sample was complete and accurate, the evaluation team and the Probation Department worked closely with one another to confirm the classification of outcomes in the validation sample data. During this process, both parties identified gaps in previous data collection processes to validate the PSA as designed. These gaps were partly due to the Department's change in case management systems during the evaluation period, which required a complex integration effort. For this study, officers looked up individuals in the old case management system to manually collect required validation data from case notes for about 200 observations. Manual data collection on this scale can introduce errors in outcomes reporting and may influence results. This limitation may be particularly influential with an overall sample size of 990 observations.

The Probation Department has already implemented data collection processes to support future validation studies. The Department's current case management system includes PSA validation-specific outcome fields for officers to collect. The Department has also implemented an automated data tracking system to remind officers to complete missing data entries promptly.

This validation study is limited in part by its relatively small sample size. To validate the PSA, the evaluation team utilized a sample of 990 observations, just one-tenth the sample size of the PSA validation completed for San Francisco County in 2021.[9] The evaluation individually examines the

---

[6] Desmarais, S. L, Singh, J. P. (2013). Risk assessment instruments validated and implemented in correctional settings in the United States. Lexington, Kentucky. *Council of State Governments, 2.* Accessed here: *https://csgjusticecenter.org/wp-content/uploads/2020/02/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf*

[7] Desmarais, S. L, Singh, J. P. (2013). Risk assessment instruments.

[8] Skog, A., Lacoe, J. (2021). Validation of the PSA in San Francisco. *California Policy Lab.*

[9] Ibid.

predictive accuracy for the six scaled scores, limiting sample sizes even further. Small sample sizes are exacerbated for demographic sub-analyses. For example, there were just 55 defendants with "other" racial or ethnic identities in the validation sample, and as few as one defendant has any one scaled score. In comparison, when the Judicial Council conducted their 2022 validation of pretrial risk assessment tools used by California Pretrial Pilot Programs, they did not perform validation analyses by racial and ethnic sub-groups if the overall sample had fewer than 1,000 observations or for any individual demographic sub-group with fewer than 200 observations.[10] Although sample sizes for Contra Costa County's validation did not reach this threshold, the evaluation had to proceed with analysis using a smaller sample size to comply with the terms of SB 36 requiring an assessment of tool predictive accuracy and bias by defendant demographic characteristics.

Additionally, pretrial outcomes were not consistently tracked throughout the evaluation period for individuals released on their own recognizance (i.e., without any Probation reporting requirements or other supervision terms). Ultimately, having comprehensive outcomes for all defendants released pretrial could have increased the sample size for this validation study. Additionally, not including outcomes from the typically low-risk own recognizance population may lead to overestimating failures.

For this validation, new criminal activity was classified as violent if the alleged offense was "strikable," meaning a serious violent felony offense that is ineligible for pretrial release. Future validation studies should consider using a broader definition in keeping with the PSA's recommended guidance of what constitutes a violent offense.

[10] Judicial Council. (2022). Pretrial risk assessment tool validation: Pretrial pilot program. Judicial Council of CA, Operations and Program Division, Criminal Justice Services.

II. Results

Compiling data with a mixed-methods approach, **the RDA study team found that most defendants released pretrial did not recidivate, and the PSA has mixed predictive accuracy in Contra Costa County**. While the PSA has fair predictive accuracy for FTA, the PSA did not have predictive accuracy for NCA or NVCA in the full sample or by racial and gender sub-groups. The tool's predictive validity may be improved with increased implementation fidelity, although mixed results may also result from small sample sizes. Due to these limitations, results should be carefully interpreted. Additionally, PSA results are just one factor the Probation Department considers when making their pretrial release recommendation to the Court. The Probation Department and the Judicial Officer of the Court (i.e., the judge) exercise their professional discretion when interpreting the PSA results, with the judge having the ultimate authority to release defendants pretrial.

This section shares study results related to local assessment practices and predictive validity findings, informing recommendations that follow.

Local Assessment Practices

RDA collected survey responses from six Probation Department Officers in the Pretrial Services Unit who are involved with administering the PSA to learn about local tool implementation. Additionally, RDA spoke with Probation administrators to understand department policies and day-to-day practices. As detailed in the preceding **Methodology** section, the evaluation team coded survey responses to identify facilitators and barriers to PSA assessment implementation and administration. The research team also incorporated descriptive statistics from the predictive validity data sample to further contextualize survey findings.

The Pretrial Services Unit demonstrated several strengths in their local administration of the PSA that support administrative fidelity and reliable scoring. All officers are satisfied with the tool, consulted required data sources and reference scoring instructions, and reported the unit has an internal quality assurance review process to ensure scoring accuracy. Further, most officers participating in the survey are confident in the tool's ability to accurately predict the risk of recidivism. To improve, the Probation Department should address scoring questions raised by survey respondents and have all unit officers participate in formal PSA training.

The County demonstrated strengths in its adherence to effective intervention in the RNR model by applying PSA results in pretrial release recommendations and supervision terms. All survey respondents use PSA results and their unit's release conditions matrix to recommend a release in their pretrial report. Ultimate authority rests with the judicial officer to determine whether an individual will be released and their supervision terms.

This report shares relevant qualitative and quantitative findings exploring local PSA assessment practices in greater detail for the following areas: (1) administration fidelity, (2) officer buy-in, and (3) release recommendations.

Administration Fidelity

Arnold Ventures designed the PSA as a simple pretrial risk assessment completed through a file review (i.e., review of justice records such as California Law Enforcement Telecommunications System (CLETS), court documents, and internal case management systems). Although the PSA is easy to use, training is required. Most respondents (67%, n = 4), but not all, have received formal training on using the PSA. In comparison, a smaller share (33%, n = 2) have participated in booster training (i.e., retraining).

Officers stated that they follow PSA administration protocols consistently and that internal practices ensure the PSA is implemented with fidelity, contributing to reliable scoring. When scoring the PSA, all respondents consulted data from existing Court and Probation computer systems, such as CLETS), court documents, and internal case management systems. Additionally, all respondents reference scoring instructions during an assessment. Each respondent also confirmed that there is a quality assurance process in place for the PSA to ensure assessment accuracy, such as supervisor review and approval.

Overall, respondents shared limited challenges related to PSA administration. When officers did report implementation barriers, they were most often associated with specific questions and their scoring. Specifically, these respondents reported PSA challenges related to an inability to reflect multiple arrests or pending charges in scoring, conceptualizing risk behaviors, and appropriately accounting for delinquent and criminal history among transitional-age youth with sealed records. Respondents also shared difficulties in handling domestic violence cases not charged with offenses that automatically disqualified them for release.

Officer Buy-In

Survey respondents expressed strong support for local use of the PSA. All officers surveyed are satisfied with the tool and are confident that a defendant will receive the same score regardless of which staff member conducts the assessment. Additionally, the majority of respondents (83%, n = 5) agree or strongly agree the PSA has accomplished the following:

- Has helped their agency establish a common language regarding defendants' risk levels;
- Is easy to read, interpret, and use;
- Increases consistency in their rules, policies, and procedures regarding assessment practices;
- Helps their agency make proper decisions regarding pretrial supervision versus custody; and
- Reduces bias in decision-making.

Most respondents further reported they are confident or completely confident that the PSA accurately captures the likelihood of a defendant missing a pretrial court hearing (i.e., FTA; 83%, n = 5), new criminal activity (i.e., NCA; 67%, n = 4), or new violent criminal activity (i.e., NVCA; 67%, n = 4).

These survey findings reflect that strong officer buy-in is essential to maintaining administrative fidelity for the PSA and ensuring reliable scoring. These results also provide insight into officers' perceptions of how well the PSA works locally before receiving results from this validation study. While most officers feel confident that the PSA tool is working accurately to predict recidivism, a slightly higher share felt confident in the tool's ability to predict an FTA than an NCA or NVCA.

Release Recommendations

In keeping with local administrative fidelity for the PSA tool and the RNR model, all survey respondents reference the release conditions matrix when determining a defendant's release level from the PSA's FTA- and NCA-scaled scores. Additionally, all respondents reference the defendant's release level and the presence of an NVCA flag to assist in making release decision recommendations for the pretrial report.

In addition to PSA results, respondents consider various factors when making release recommendations in the pretrial report. These factors include the following:

- Details of the current (i.e., instant) offense;

- Criminal history;
- Current community supervision;
- Status of pending case(s);
- Potential danger to victims and the community; and
- History of failing to appear in Court.

When developing release recommendations, respondents prioritize these factors differently. Some respondents weigh all factors equally, while others prioritize public safety, instant offense details, PSA results, and victim statements. Key considerations include victim and community safety, prior criminal record, and new charges or failures to appear. As mentioned previously in the **Introduction**, the ultimate release decision authority in Contra Costa County rests with the Judicial Officer of the Court (i.e., the judge), who can deny pretrial release, release defendants on their own recognizance, or release defendants under the supervision of Probation's Pretrial Services Unit. The PSA is one factor that is used to by the Probation Department to help inform their recommendations to the Court and one factor in the ultimate decision made by the Judicial Officer.

Predictive Validity

To examine the internal reliability and predictive accuracy of the PSA in Contra Costa County, the RDA evaluation team normed the instrument on the local population and executed several predictive validity analyses to measure the extent to which the tool accurately predicts FTA, NCA, and NVCA. These findings establish the distribution of assessment scores for the evaluation period and whether PSA items are theoretically relevant and connected with the measurement of criminogenic risk of local defendants. As described in the preceding **Methodology** section, these analyses used administrative data from Contra Costa County related to defendants who have received assessments, their assessment scores, and outcomes.

About 54% of all defendants assessed before January 1, 2024, were ultimately released pretrial, including 55% male and 52% female defendants. Concerning race/ethnicity, 56% of Black/African American defendants, 55% of Hispanic/Latino defendants, 52% of White defendants, and 50% of defendants of an "Other" race were released pretrial. **Differences in pretrial release rates did not rise to the level of statistical significance for gender or race, meaning that observed differences may be due to chance.** See **Appendix B** for a detailed breakdown of defendant characteristics by release decision, including PSA scaled scores.

Most released defendants in the predictive validity sample did not have an FTA (73%, n = 719), NCA (91%, n = 901), or NVCA (99%, n = 983). While the PSA has fair predictive accuracy for FTA across the full validation sample and defendants in different demographic groups, the PSA has weak predictive accuracy for NCA or NVCA. As discussed in the **Limitations** section, it is unclear whether predictive accuracy problems stem from the small number of released individuals that had an NCA (n = 89) or NVCA (n = 7), outcome data collection limitations, or issues with the relevance of PSA's risk factors in Contra Costa County. Due to these significant limitations, results should be carefully interpreted. Additionally, it is essential to note that PSA results are just one factor the Probation Department considers when making their pretrial release recommendation to the Court. The Probation Department and the Judicial Officer of the Court (i.e., the judge) exercise their professional discretion when interpreting the PSA results, with the judge having the ultimate authority to release defendants pretrial.

This section reviews complete internal reliability and predictive validity findings in the following structure: 1) norms & internal reliability, 2) full sample predictive validity, and 3) race/ethnicity and gender predictive validity.

Norms & Internal Reliability

RDA analyzed PSA data from a sample of 1,114 defendants in Contra Costa County to assess the norms and internal reliability of the tool. The analysis examined key PSA scaled scores, including FTA and NCA scores and the NVCA flag, across the full sample and subgroups, including gender and race/ethnicity.

As seen from **Table 3**, PSA scaled scores for defendants in Contra Costa County revealed several significant differences based on demographic subgroups. On average, defendants scored a 3.2 for FTA and 3.4 for NCA. Average scaled scores varied across gender and race, with male defendants yielding an average FTA scaled score of 3.1 and an NCA scaled score of 3.4 and females scoring slightly higher, with an average FTA scaled score of 3.3 and an NCA scaled score of 3.2. These gender differences were not significantly different, however.

White defendants (FTA: 3.6, NCA: 3.7) had significantly higher average scaled scores than Hispanic defendants (FTA: 3.0, NCA: 3.1), Black defendants (FTA: 3.1, NCA: 3.5), and defendants of an "Other" race or ethnicity (FTA: 2.8, NCA: 3.1). These differences in scores across racial groups were statistically significant as indicated with the asterisks in **Table 3**, meaning that it is unlikely that these differences are due to chance alone.

Table 3. PSA Average Scaled Score for Overall Sample & Subgroups

	Overall (*N* = 1,114)	Male (*n* = 930)	Female (*n* = 182)	White (*n* = 293)	Hispanic (*n* = 304)	Black (*n* = 457)	"Other" (*n* = 60)
Average Scaled Score							
FTA	3.2	3.2	3.3	3.6***	3.0***	3.1***	2.8***
NCA	3.4	3.4*	3.2*	3.7***	3.1***	3.5***	3.1***
NVCA	.0	.0	.0	.0	.0	.0	.0

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

Note: Gender was missing for two individuals, resulting in an N of 1,112.

The PSA's internal reliability, measured by Cronbach's alpha, was below the commonly accepted threshold of .70 for all measures and subgroups (see **Table 4**). The FTA score showed an alpha of .38 for the overall sample, indicating low internal consistency. The NCA score had an alpha of .61, closer to the acceptable range but still below the ideal threshold of .70. The NVCA flag had the lowest alpha at .23, suggesting weak internal reliability for this measure across all subgroups.

Table 4. PSA Internal Reliability Coefficients for Overall Sample & Subgroups

	Overall (*N* = 1,114)	Male (*n* = 930)	Female (*n* = 182)	White (*n* = 293)	Hispanic (*n* = 304)	Black (*n* = 457)	"Other" (*n* = 60)
Alpha							
FTA	.38	.37	.44	.32	.40	.39	.42
NCA	.61	.60	.62	.54	.61	.62	.60
NVCA	.23	.25	.22	.21	.20	.23	.39

PSA scaled scores were further broken down by levels (see **Tables 5, 6,** and **7**), indicating that a significant portion of the sample fell into higher risk categories (i.e., scaled scores of three or more) for both FTA and NCA. Notably, White defendants were assessed as higher risk at a greater rate than other racial/ethnic groups for both predicted outcomes.

Table 5. PSA Distribution for Overall Sample & Subgroups by FTA Scaled Score

	Overall (*N* = 1,114)		Male* (*n* = 930)		Female* (*n* = 182)		White*** (*n* = 293)		Hispanic*** (*n* = 304)		Black*** (*n* = 457)		"Other"*** (*n* = 60)	
FTA Score	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
1	180	16.2	147	15.8	33	18.1	28	9.6	69	22.7	72	15.8	11	18.3
2	194	17.4	168	18.1	26	14.3	36	12.3	50	16.4	91	19.9	17	28.3
3	283	25.4	234	25.2	48	26.4	76	25.9	73	24.0	120	26.3	14	23.3
4	233	20.9	208	22.4	25	13.7	78	26.6	57	18.8	90	19.7	8	13.3
5	162	14.5	125	13.4	36	19.8	49	16.7	41	13.5	63	13.8	9	15.0
6	62	5.6	48	5.2	14	7.7	26	8.9	14	4.6	21	4.6	1	1.7

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

Table 6. PSA Distribution for Overall Sample & Subgroups by NCA Scaled Score

	Overall (*N* = 1,114)		Male (*n* = 930)		Female (*n* = 182)		White*** (*n* = 293)		Hispanic*** (*n* = 304)		Black*** (*n* = 457)		"Other"*** (*n* = 60)	
NCA Score	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
1	102	9.2	77	8.3	25	13.7	22	7.5	36	11.8	36	7.9	8	13.3
2	143	12.8	122	13.1	21	11.5	21	7.2	58	19.1	57	12.5	7	11.7
3	354	31.8	291	31.3	62	34.1	90	30.7	100	32.9	139	30.4	25	41.7
4	305	27.4	261	28.1	44	24.2	90	30.7	65	21.4	138	30.2	12	20.0
5	135	12.1	114	12.3	21	11.5	44	15.0	26	8.6	58	12.7	7	11.7
6	75	6.7	65	7.0	9	4.9	26	8.9	19	6.3	29	6.3	1	1.7

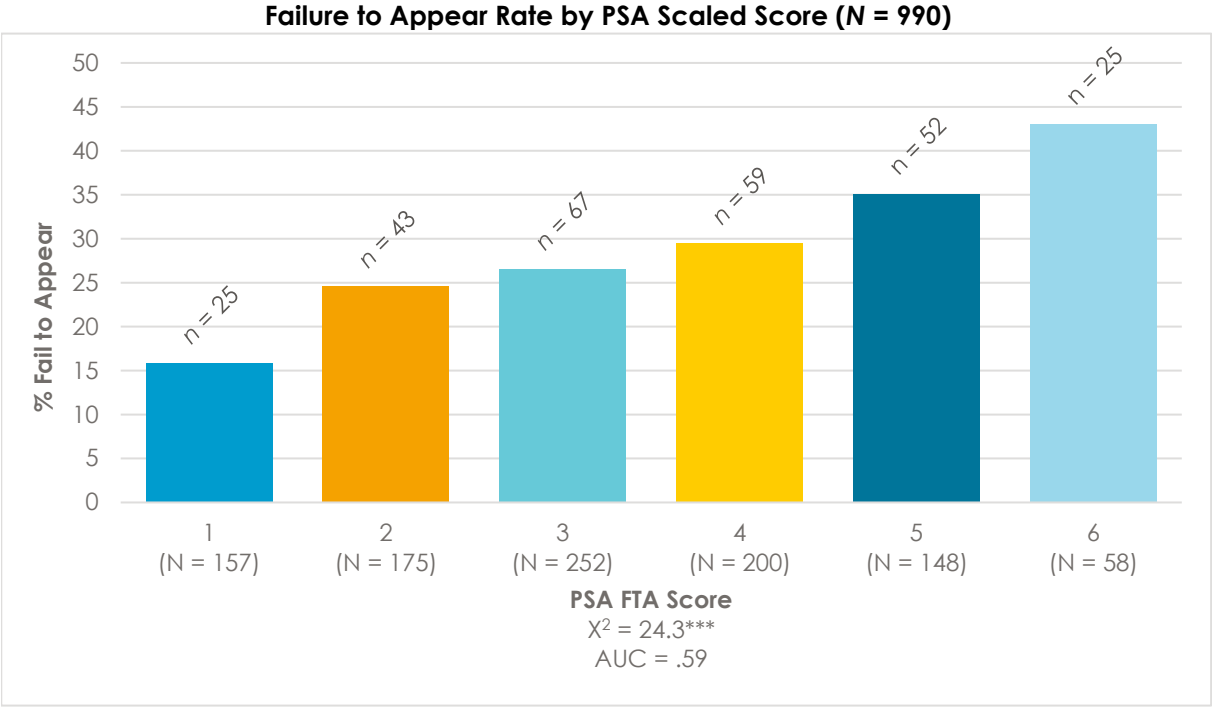
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 7. PSA Distribution for Overall Sample & Subgroups by NVCA Flag

	Overall (*N* = 1,114)		Male (*n* = 930)		Female (*n* = 182)		White (*n* = 293)		Hispanic (*n* = 304)		Black (*n* = 457)		"Other" (*n* = 60)	
NVCA Flag	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
No	1,082	97.1	903	97.1	178	97.8	281	95.9	300	98.7	441	96.5	60	100.0
Yes	32	2.9	27	2.9	4	2.2	12	4.1	4	1.3	16	3.5	0	.0

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

The findings from these norms and internal reliability analyses indicate statistically significant differences in PSA scaled scores across racial groups, particularly between White defendants and other racial/ethnic groups. However, the internal reliability of the PSA for all three predicted outcomes (i.e., FTA, NCA, and NVCA) falls below the recommended .70 threshold, suggesting the need for caution when interpreting these scores in decision-making processes.



* p < .05, ** p < .01, *** p < .001
n values reflect the number of defendants who failed in each category.

Full Sample Predictive Validity

Table 8 and **Figure 1** present the FTA rate by PSA scaled score based on a sample size of 990 defendants in Contra Costa County. The findings demonstrate a clear trend as the PSA scaled score the FTA rate increase. For those with a PSA scaled score of 1, the FTA rate is 16%, with 25 out of 157 defendants failing to appear. This rate rises to 25% for those with a scaled score of 2, with 43 out of 175 defendants failing to appear. Similarly, 27% of defendants with a PSA scaled score of 3 failed to appear (n = 67), while 30% of those with a scaled score of 4 (n = 59) did not. The FTA rate continues to increase for individuals with PSA scaled scores of 5 and 6, reaching 35% (n = 52) and 43% (n = 25), respectively.

The statistical analysis supports the significance of these findings. A Chi-square test reveals a strong relationship between PSA scaled score and FTA rate, with a value of $X^2 = 24.3$ and a p-value less than .001, indicating that the association is statistically significant and not due to chance. Furthermore, this model's AUC is .59, suggesting that the PSA scaled score offers fair accuracy in predicting the likelihood of failing to appear in Court.

Finally, logistic regression results measuring the relationship between the PSA scaled score and the likelihood of failing to appear in Court demonstrate a statistically significant relationship between the PSA scaled score and the likelihood of an FTA. Specifically, the marginal effect results indicate

Table 8. Failure to Appear Rate by PSA Scaled Score (N = 990)[a]

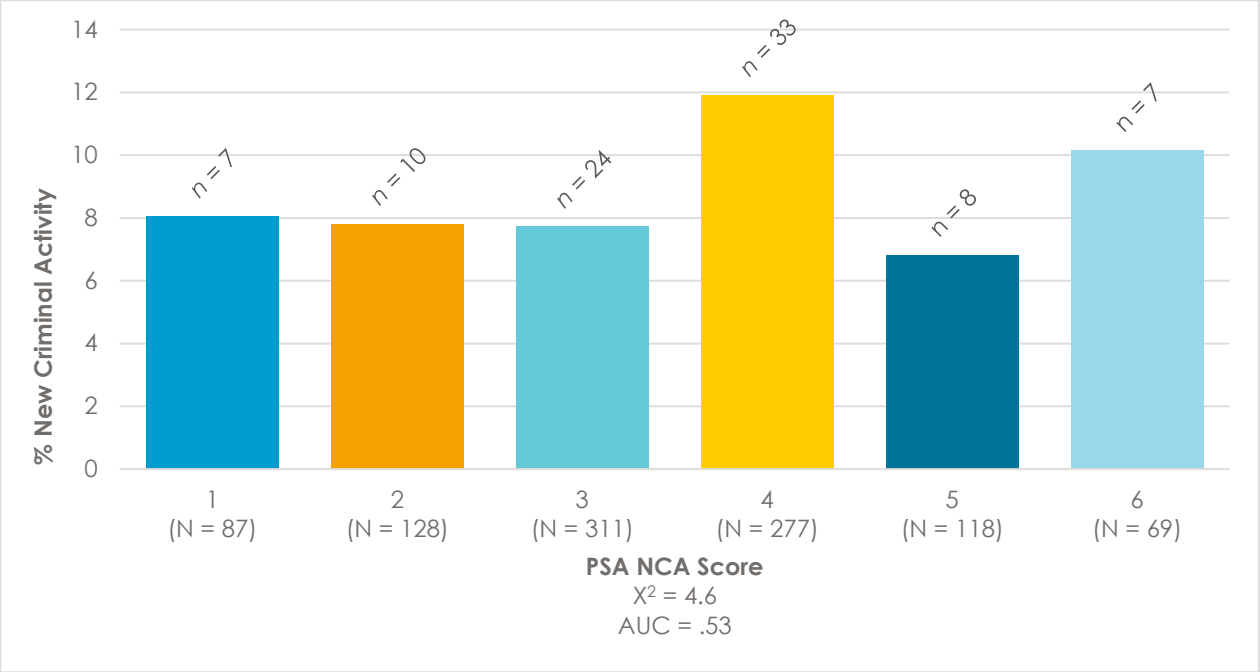
	n	Failure rate (%)
PSA Scaled Score		
1	25	15.9
2	43	24.6
3	67	26.6
4	59	29.5
5	52	35.1
6	25	43.1
Total	*271*	*27.4*
	$\chi^2 = 23.3^{***}$	
AUC-ROC, PSA scaled score	.59	
Lower 95% CI	.55	
Upper 95% CI	.63	
Pearson's r	$.15^{***}$	
Logistic Regression	$.04^{***}$	

Note. PSA = Public Safety Assessment; AUC-ROC = area under the receiver operating characteristic curve; CI = confidence interval; Logistic regression results present marginal effects from a binomial logit model.
* $p < .05$, ** $p < .01$, *** $p < .001$
[a]. Sample totals: 1 = 157, 2 = 175, 3 = 252, 4 = 200, 5 = 148, 6 = 58, Total = 990

that a one-point increase in the PSA scaled score is associated, on average, with a four-percentage point increase in the likelihood of an FTA for released defendants (see **Appendix D** for complete logistic regression results).

Table 9 and **Figure 2** illustrate the rate of new criminal activity by PSA scaled score in the sample of 985 defendants from Contra Costa County. The findings indicate a varied pattern in new criminal activity rates across PSA scaled scores. For individuals with a PSA scaled score of 1, the new criminal activity rate is 8.1% (n = 7). The rate remains relatively stable for PSA scaled scores of 2 and 3, with 7.8% (n = 10) and 7.7% (n = 24), respectively, engaging in new criminal activity. However, individuals with a PSA scaled score of 4 exhibit a notably higher rate of new criminal activity, at 11.9% (n = 33). For those with a PSA scaled score of 5, the rate decreases to 6.8% (n =8), while individuals with a PSA scaled score of 6 show a slight increase in new criminal activity, with a rate of 10.1% (n = 7).

New criminal activity Rate by PSA Scaled Score (*N* = 990)



* p < .05, ** p < .01, *** p < .001
n values reflect the number of defendants who failed in each category.

Table 9. New Criminal Activity Rate by PSA Scaled Score (N = 990)[a]

	n	**Failure rate (%)**
PSA Scaled Score		
1	7	8.1
2	10	7.8
3	24	7.7
4	33	11.9
5	8	6.8
6	7	10.1
Total	*89*	*9.0*
	$X^2 = 4.6$	
AUC-ROC, PSA scaled score	.53	
Lower 95% CI	.47	
Upper 95% CI	.59	
Pearson's r	.02	
Logistic Regression	.00	

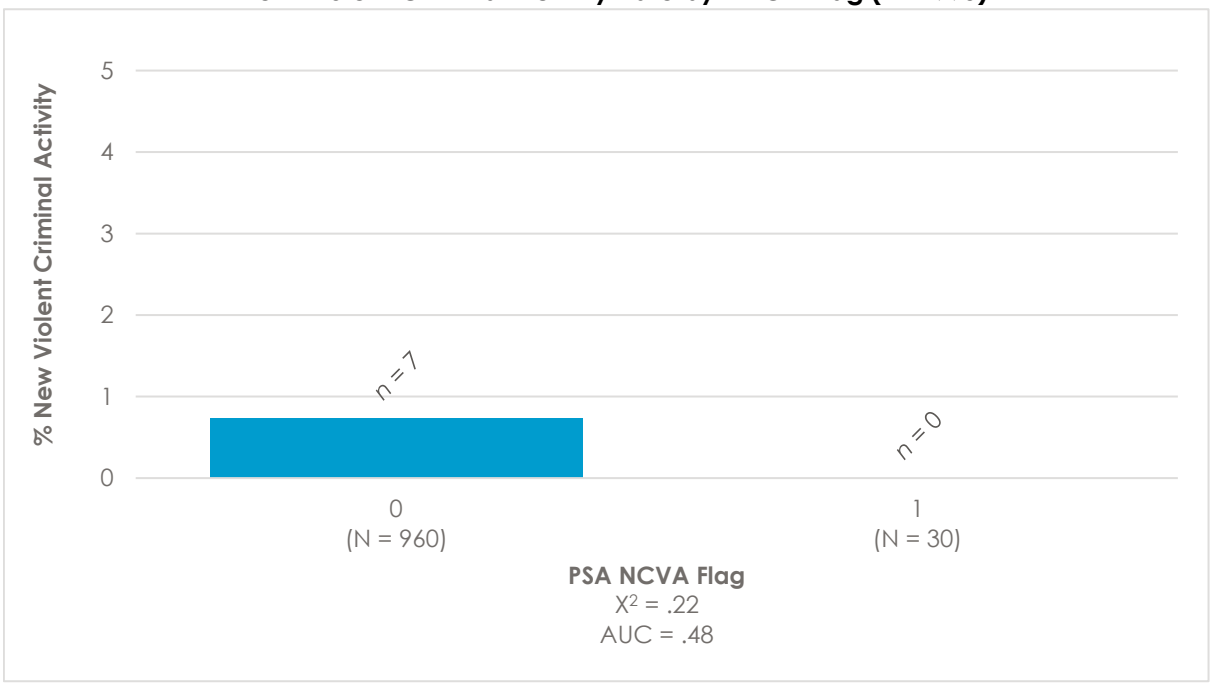
Note. PSA = Public Safety Assessment; AUC-ROC = area under the receiver operating characteristic curve; CI = confidence interval; Logistic regression results present marginal effects from a binomial logit model.
* *p* < .05, ** *p* < .01, *** *p* < .001
[a.] Sample totals: 1 = 87, 2 = 128, 3 = 311, 4 = 277, 5 = 118, 6 = 69, Total = 990

Statistical analysis provides additional insights. The Chi-square test ($X^2 = 4.6$) is not as strong as the failure to appear analysis, suggesting a weaker relationship between PSA scaled scores and new criminal activity rates. Additionally, this model's AUC is 0.53, indicating less than fair, or slight, predictive power in using PSA scaled scores to forecast new criminal activity. In short, the relationship between PSA scaled scores and new criminal activity is not as predictive as PSA scaled scores and failure to appear. While individuals with a PSA scaled score of 4 demonstrate a higher likelihood of engaging in new criminal activity, overall, the PSA scaled score shows only moderate effectiveness in predicting such outcomes. The logistic regression result indicating a marginal effect of 0.00 between the PSA scaled score and the likelihood of committing new criminal activity also means that, on average, a one-unit increase in the PSA scaled score does not have a meaningful effect on the probability of engaging in new criminal activity.

Figure 3 and **Table 10** present the rate of NVCAs based on the PSA NVCA flag for the sample of 990 defendants. Nearly all defendants in the sample (n = 960) were assigned a low-risk NVCA flag (i.e., a value of 0), while only 30 defendants were flagged as high-risk (i.e., a value of 1). Among defendants classified as low risk, 0.73% (n = 7) committed a new violent offense. In contrast, none of the individuals flagged as high risk committed a new violent offense, resulting in a 0% failure rate for this group.

Statistical analysis provides additional context to these findings. The Chi-square test ($X^2 = .22$, $p = .64$) indicates no significant association between the PSA NVCA flag and the occurrence of new violent criminal activity. Additionally, the AUC for this model is .48, which suggests poor predictive accuracy; an AUC of .5 would indicate no predictive power, and the result here is below that threshold. The Pearson correlation coefficient (r = -.02) also shows no meaningful correlation

New Violent Criminal Activity Rate by NVCA Flag (*N* = 990)



* p < .05, ** p < .01, *** p < .001
n values reflect the number of defendants who failed in each category.

Table 10. New Violent Criminal Activity Rate by NVCA Flag (N = 990)[a]

	n	Failure rate (%)
PSA Flag		
0	7	.73
1	0	.00
Total	7	.71
	χ^2 = .22	
AUC-ROC, PSA Flag	.48	
Lower 95% CI	.48	
Upper 95% CI	.49	
Pearson's r	-.02	
Logistic Regression	.00	

Note. PSA = Public Safety Assessment; AUC-ROC = area under the receiver operating characteristic curve; CI = confidence interval; Logistic regression results present marginal effects from a binomial logit model.
* $p < .05$, ** $p < .01$, *** $p < .001$
[a.] Sample totals: 0 = 960, 1 = 30, Total = 990

between the NVCA flag and actual violent offense outcomes. Finally, the logistic regression model yields a marginal effect of .00, indicating that the PSA NVCA flag has no meaningful impact on predicting new violent criminal activity.

Incorporating methods utilized in the California Policy Lab's San Francisco County PSA validation, the evaluation team also examined correct risk classification on a binary scale.[11] Specifically, the evaluation team categorized PSA scaled scores greater than three as high-risk to assess the PSA's ability to accurately predict failures and non-failures with additional measures such as negative and positive predictive value.[12]

As demonstrated in **Figure 4**, there are four possible classifications from a set of two possible predictions and two outcomes: true failures (i.e., a failure outcome predicted as a failure), false failures (i.e., a non-failure outcome incorrectly predicted as a failure), true non-failures (i.e., a non-failure outcome predicted as a non-failure), and false non-failures (i.e., a failure outcome incorrectly predicted as a non-failure).

Figure 4. Classification of Predicted & Actual Pretrial Release Failure and Non-Failure Outcomes

	Predicted Failure	**Predicted Non-Failure**
Failure Outcome	True Failure	False Non-Failure
Non-Failure Outcome	False Failure	True Non-Failure

[11] Skog, A., Lacoe, J. (2021). Validation of the PSA in SF.
[12] These are the same scaled score cut-offs the California Policy Lab used in the primary analysis for the San Francisco County PSA validation.

Table 11. Classification, by Predicted Outcome

	High-Risk	%	Accuracy (Correct Classification True Failures & Non-Failures)	Positive Predictive Value (Precision of Failure Prediction)	False Negative Rate (Missed Fail)	Negative Predictive Value (Precision of No Failure Prediction)	False Positive Rate (False Alarm)
FTA	406	41%	59%	33%	38%	77%	50%
NCA	464	47%	54%	10%	46%	92%	46%
NVCA	30	3%	96%	0%	3%	99%	100%

Note: Defendants classified as "high risk" for NVCA had an NVCA flag of "yes."

Table 11 shows that the FTA has the highest positive predictive value across the three PSA outcomes. This finding aligns with the validation results discussed in previous paragraphs, showing that FTA scaled scores had the strongest association with an FTA outcome. Ultimately, the NVCA flag had no positive predictive value because none of the defendants with an NVCA flag ultimately failed.[13] As a result, the NVCA flag also had a 100% false positive rate.[14] The NVCA flag additionally had the highest negative predictive value (99%), meaning almost all predictions of no failure were accurate.[15] The share of false negatives, i.e., missed failures, was also the smallest (3%) for the NVCA flag.[16] Based on the NVCA flag's high share of true non-failure, the NVCA has the highest accuracy overall.[17]

For this validation study, we are most concerned with accurately predicting the likelihood that any of the predicted pretrial failure outcomes will happen to account for any risks to public safety properly. These statistics are captured in the positive predictive value and false negative rate (i.e., missed fail rate). All positive predictive values were ultimately low, while NCA's false negative rate reached 46%.

[13] Positive predictive value is calculated by dividing true failures by all predicted failures.
[14] The false positive rate is calculated by dividing false failures by all non-failures; it represents the share of non-failures that were incorrectly predicted to fail (i.e., false alarm rate).
[15] Negative predictive value is calculated by dividing true non-failures by all predicted non-failures.
[16] The false negative rate is calculated by dividing false non-failures by all failures; it represents the share of failures that were not predicted (i.e., missed fail rate).
[17] The accuracy rate is calculated as the sum of true failures and non-failures divided by the total population of failures and non-failures.

In addition to exploring the relationship between PSA scaled scores and predicted outcomes (i.e., FTA, NCA, and NVCA), the evaluation team investigated the relationship between individual risk factors and their related outcomes. Replicating analysis performed by the California Policy Lab's PSA validation, the RDA evaluation team investigated the failure rate observed for each risk factor in addition to performing bivariate and multivariate regression analyses.[18] **Table 12** presents the risk factor's name, item number, and how each risk factor is scored. Each predicted outcome has two columns that display the frequency of assessed defendants in each risk factor scoring bracket and their failure rate. The FTA, NCA, and NVCA rates for each risk factor should increase as the

Table 12. Individual Risk Factors & Predicted Outcomes

Risk Factor			Score	N	FTA Rate	N	NCA Rate	N	NVCA Rate
Item 1	Age at current arrest	23 or older	0			840	**9.4%**		
		22 or younger	2			150	**7.3%**		
Item 2	Current violent offense	No	0					929	**0.8%**
		Yes	2					61	**0.0%**
Item 2A	Current violent offense and 20 years old or younger	No	0					984	**0.7%**
		Yes	1					6	**0.0%**
Item 3	Pending charge at the time of the arrest	No	0	769	26.3%	769	**9.1%**	769	**0.8%**
		Yes	1^	221	31.2%	221	**8.6%**	221	**0.5%**
Item 4	Prior misdemeanor conviction	No	0			336	6.8%		
		Yes	1			654	10.1%		
Item 5	Prior felony conviction	No	0			474	7.8%		
		Yes	1			516	10.1%		
Item 5A	Prior conviction (misdemeanor or felony)	No	0	261	20.7%			261	**0.8%**
		Yes	1	729	29.8%			729	**0.7%**
Item 6	Prior violent conviction	No	0			782	**8.8%**	782	**0.9%**
		Yes, 1-2	1			190	**10.0%**	190	**0.0%**
		Yes, 3+	2			18	**5.6%**	18	**0.0%**
Item 7	Prior failure to appear in the past 2 years	No	0	587	23.2%	587	**9.0%**		
		Yes, 1	2^	217	28.6%	217	**7.8%**		
		Yes, 2+	4^	185	39.5%	185	**10.3%**		
Item 8	Prior failure to appear older than 2 years	No	0	495	23.2%				
		Yes	1	495	31.5%				
Item 9	Prior sentence to incarceration	No	0			298	6.7%		
		Yes	2			690	10.0%		

Note: ^ Item 3 score is 0 for "No" responses across all three outcomes, while a "Yes" response is a 1 for FTA and NVCA and a 3 for NCA. Item 7 score is 0 for "No" responses across FTA and NCA, while a "Yes, 1" is a 1 for NCA and "Yes, 2+" is a 2 for NCA.

[18] Skog, A., Lacoe, J. (2021). Validation of the PSA in SF. See **Appendix E** and **F** for complete reporting on the bivariate and multivariate binomial logit model results.

score rises and be a statistically significant predictor of failure, individually and in combination with other relevant risk factors.

In the case of FTA outcomes, the FTA rate does increase as the score rises for every FTA risk factor. Additionally, items 5a, 7, and 8 have individual statistically significant relationships with an FTA outcome, meaning those risk factors have predictive value for FTA outcomes that are not likely due to chance.[19] However, only item 7 (i.e., prior failure to appear in the past 2 years) and item 8 (i.e., prior failure to appear older than 2 years) had a statistically significant relationship with an FTA outcome when all FTA risk factors were tested together, meaning all other items were less predictive of an FTA overall.

The NCA and NVCA rates that do not conform with our expectations are identified in red text in Table 12, indicating that the risk factors do not predict outcomes well. The NCA rate decreases as scores increase in four of seven NCA risk factors, while the NVCA rate decreases as scores increase for all five NVCA risk factors.[20] Overall, the NCA and NVCA rates are low and vary by less than one or two percentage points in most cases.

The three predicted outcomes are included side by side in **Table 12** to facilitate identifying trends in issues with items used across multiple outcomes. Items that did not perform as expected for at least two predicted outcomes are enclosed with a red border. Specifically, **Table 12** identifies issues with item 3, "pending charge at the time of arrest," and item 6, "prior violent convictions." Of note, although the FTA rate does increase with item 3 scoring, the item is not a statistically significant predictor of that outcome.

While the evaluation team cannot offer scoring changes to the PSA to enhance predictive accuracy based on these results alone, identifying factors that are not operating as expected may provide direction for specific follow-up training with officers in the unit in addition to efforts already underway, such as data tracking systems to alert officers to complete data entry. Officers participating in the survey portion of this validation study raised challenges with their inability to reflect multiple arrests or pending charges in scoring item 3; however, this specific scoring issue does not explain the problems with item 3 highlighted in **Table 12**. Defendants with at least one charge had an NCA or NVCA at a lower rate than those with none.

[19] Statistical significance and coefficients are reported in **Appendix E** and **F** and were measured using a bivariate binomial logit model that assessed the relationship between individual risk factors and the predicted outcome, as well as a multivariate binomial logit model that evaluated the relationship of all relevant risk factors with the predicted outcome.
[20] None of the NCA or NVCA risk factors had a statistically significant relationship with the predicted outcome in a bivariate logit model or together in a multivariate logit model. See **Appendix E** and **F** for complete reporting on the bivariate and multivariate binomial logit model results.

Race & Gender Predictive Validity

The preceding predictive validity analyses were calculated again across race and gender to determine if the PSA correctly assesses risk levels for different sub-groups and detect any bias in tool results. The RDA evaluation team did not detect serious bias concerns by different defendant identities. Rates of FTA increased as scaled scores rose for every sub-analysis and produced fair AUC values, indicating the tool performs with similar predictive accuracy regardless of race or gender. NCA scaled scores and the NVCA flag were relatively weak predictors of their respective outcomes across each defendant identity, mirroring PSA's limited predictive accuracy for those outcomes across the overall sample as opposed to identifying bias concerns. These findings should be interpreted cautiously because of the relatively small overall sample size (i.e., 990 defendants) and even smaller sub-analysis sample sizes by both PSA scaled scores and demographic variables. For the NCA and NVCA sub-analyses, the sample size of defendants in any one sub-group for a scaled score level can be as low as four. Mentioned previously, the Judicial Council's 2022 validation study did not evaluate tool performance by racial and ethnic sub-groups if the overall sample had fewer than 1,000 observations or for any individual demographic sub-group with fewer than 200 observations.[21] Ultimately, demographic sub-analyses are included in this study despite not reaching the Judicial Council observation thresholds to comply with the terms of SB 36.

Predictive Validity by Race

Figure 5 presents FTA rates by PSA scaled score and race/ethnicity, based on the sample of 990 defendants from Contra Costa County. This analysis allows for an understanding of how FTA rates differ across racial/ethnic subgroups and PSA scaled scores.

For White defendants (*N* = 265), the FTA rate rises steadily as the PSA scaled score increases. The FTA rate is 18% for those with a PSA scaled score of 1 and reaches 48% for defendants with a PSA scaled score of 6. The Chi-square test does not show a significant relationship between PSA scaled score and FTA for White defendants ($X^2 = 8.6$, $p > .05$), although the AUC of 0.58 indicates fair predictive power for PSA scaled scores in this group.

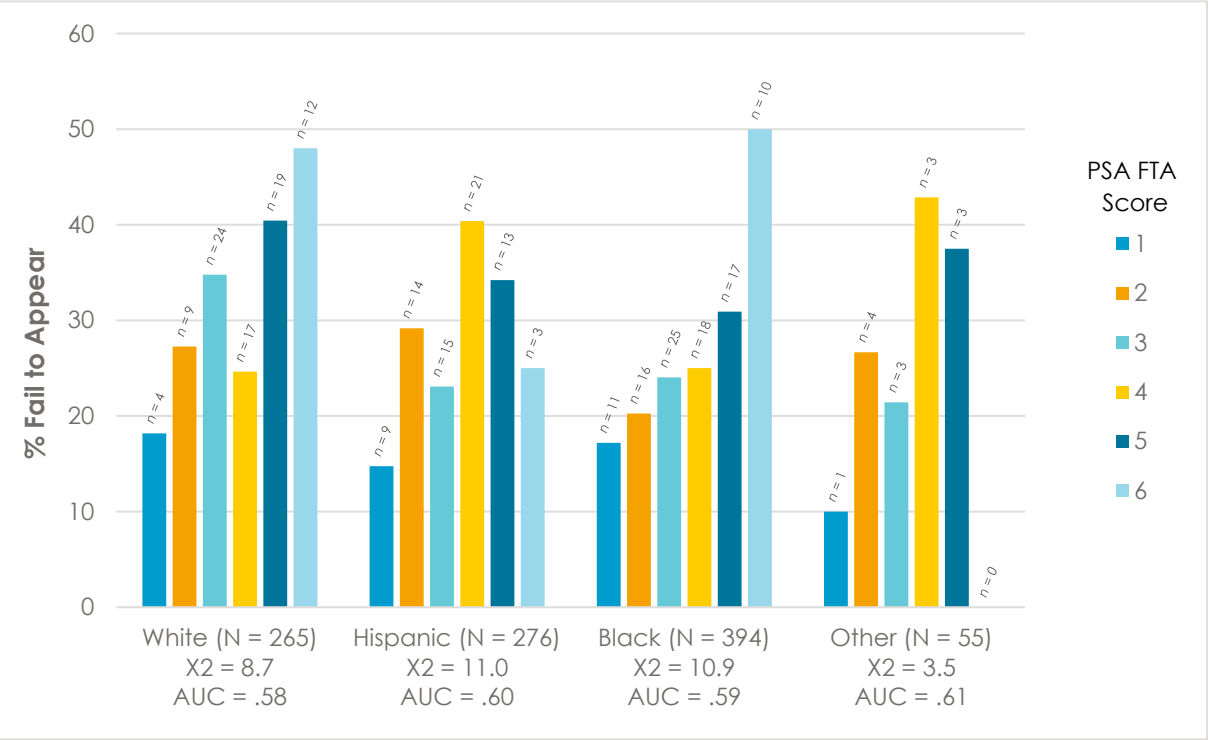
The trend is similar among Hispanic defendants (*N* = 276), with an irregular increase in FTA rates as PSA scaled scores rise. For those with a PSA scaled score of 1, the FTA rate is 15%, increasing to 25% for defendants with a PSA scaled score 6. The Chi-square test is not statistically significant ($X^2 = 11.0$, $p > .05$), with an AUC of 0.60, suggesting fair, slightly better predictive accuracy compared to White defendants.

For Black defendants (*N* = 394), the FTA rates clearly increase with higher PSA scaled scores. The FTA rate starts at 17% for those with a PSA scaled score of 1 and reaches 50% for those with a PSA scaled score of 6. The Chi-square test ($X^2 = 10.9$, $p > .05$) indicates there is not a statistically significant association between PSA scaled score and FTA in this group, and the AUC is 0.59, indicating fair predictive validity of the PSA for Black individuals.

For defendants categorized as an "Other" racial or ethnic identity (*N* = 55), the FTA rates again increase irregularly with the PSA scaled score. However, the sample size is the smallest across defendant groups. The FTA rate is 10% for a PSA scaled score of 1 and rises to 43% for those with a PSA scaled score of 4. The Chi-square test ($X^2 = 3.9$, $p > .05$) is not statistically significant, while the AUC is 0.61, indicating fair to good predictive capacity for this group.

[21] Judicial Council. (2022). Pretrial risk assessment tool validation.

Figure 5. Failure to Appear Rate by PSA Scaled Score & Race



* p < .05, ** p < .01, *** p < .001
n values reflect the number of defendants who failed in each category.

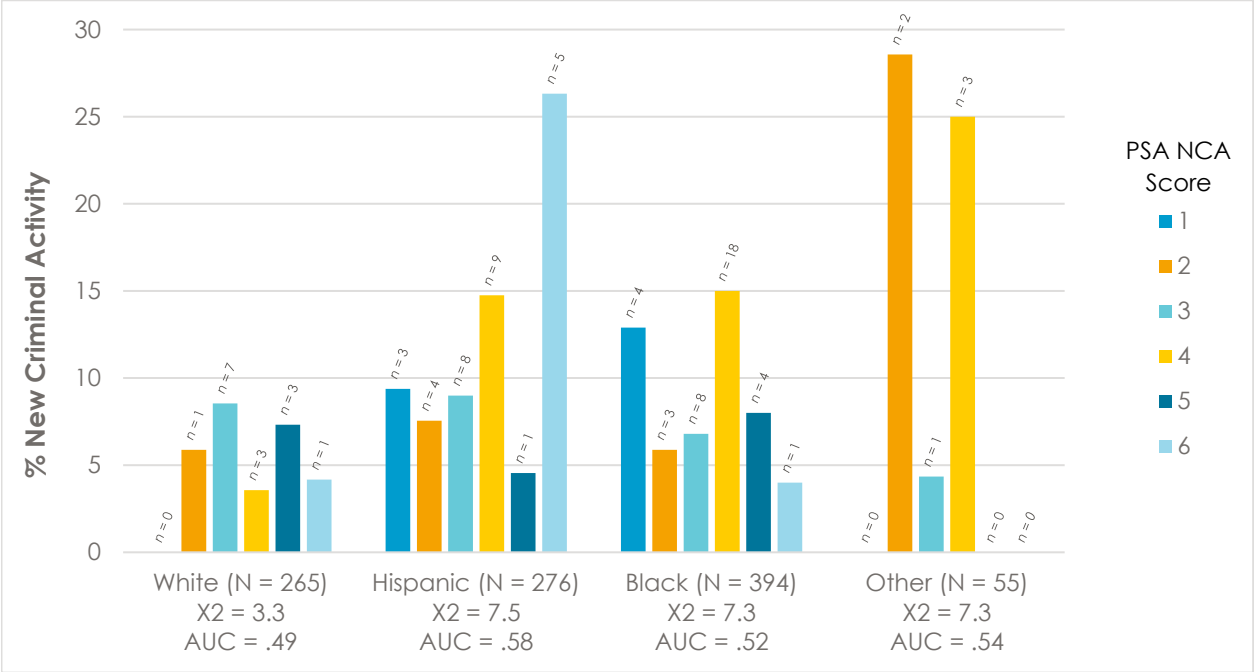
In summary, across all racial and ethnic subgroups, FTA rates generally increase with higher PSA scaled scores, demonstrating that PSA scaled scores are a valuable predictor of failure to appear regardless of race or ethnicity. The Chi-square values for Hispanic and Black defendants are the highest, albeit not statistically significant. The AUC values, which range from 0.58 to 0.61, indicate fair to good predictive accuracy of the PSA scaled score in assessing FTA risk across all subgroups.

Figure 6 shows the rate of new criminal activity by PSA scaled score and race/ethnicity for the sample of interest. This analysis allows for an understanding of how NCA rates differ across racial/ethnic subgroups with PSA scaled scores.

For White defendants (*N* = 265), the rate of new criminal activity is relatively low across all PSA scaled scores. The NCA rate begins at 0% for PSA scaled scores of 1 and peaking at 9% for PSA scaled score 3. The Chi-square test ($X^2 = 3.3$, $p > .05$) suggests no significant relationship between PSA scaled score and NCA in this group, and the AUC of 0.49 indicates poor predictive accuracy of the PSA scaled score for new criminal activity among White defendants.

There is a stronger visual relationship among Hispanic defendants (N = 276) between PSA scaled scores and new criminal activity, although the NCA rate increases inconsistently. The NCA rate is 9% for those with a PSA scaled score of 1 and increases to 26% for those with a PSA scaled score

Figure 6. New Criminal Activity Rate by PSA Scaled Score & Race



* p < .05, ** p < .01, *** p < .001

n values reflect the number of defendants who failed in each category.

of 6. The Chi-square test ($X^2 = 7.5$, $p > .05$) indicates no statistically significant relationship between PSA scaled score and NCA for Hispanic defendants, and the AUC of 0.58 shows fair predictive accuracy.

For Black defendants ($N = 394$), the NCA rate does not follow a clear pattern, with NCA rates starting at 13% for PSA scaled score of 1 and peaking at 15% for PSA scaled score of 4 before declining to 4% for PSA scaled score 6. The Chi-square test ($X^2 = 7.3$, $p > .05$) confirms the visually inconsistent pattern. It reveals no statistically significant association between PSA scaled score and NCA for Black defendants, while the AUC of 0.52 indicates less than fair, or slight predictive accuracy.

For defendants categorized as an "Other" racial or ethnic identity ($N = 55$), there is a pronounced increase in new criminal activity rates at two scaled scores. In contrast, the majority of other NCA rates are 0%. The NCA rate is 0% for those with a PSA scaled score of 1, rising dramatically to 12% and 25% for PSA scaled scores of 2 and 4, respectively. The Chi-square test ($X^2 = 7.3$, $p > .05$) is insignificant, indicating no statistically significant relationship between PSA scaled score and NCA for Hispanic defendants. However, the AUC of 0.54 suggests fair predictive power for this subgroup.

The relationship between PSA scaled scores and new criminal activity varies across racial and ethnic subgroups. Ultimately, Chi-square results indicated the relationship between scaled scores and NCA was statistically insignificant for each racial and ethnic group examined. Hispanic defendants and defendants of an "Other" race or ethnicity had stronger associations between PSA scaled scores and NCA rates, as evidenced by higher AUC scores. The PSA scaled score for White defendants appears to be a weak predictor of new criminal activity, with the lowest AUC. The AUC values do not support strong predictive validity when broken down by race.

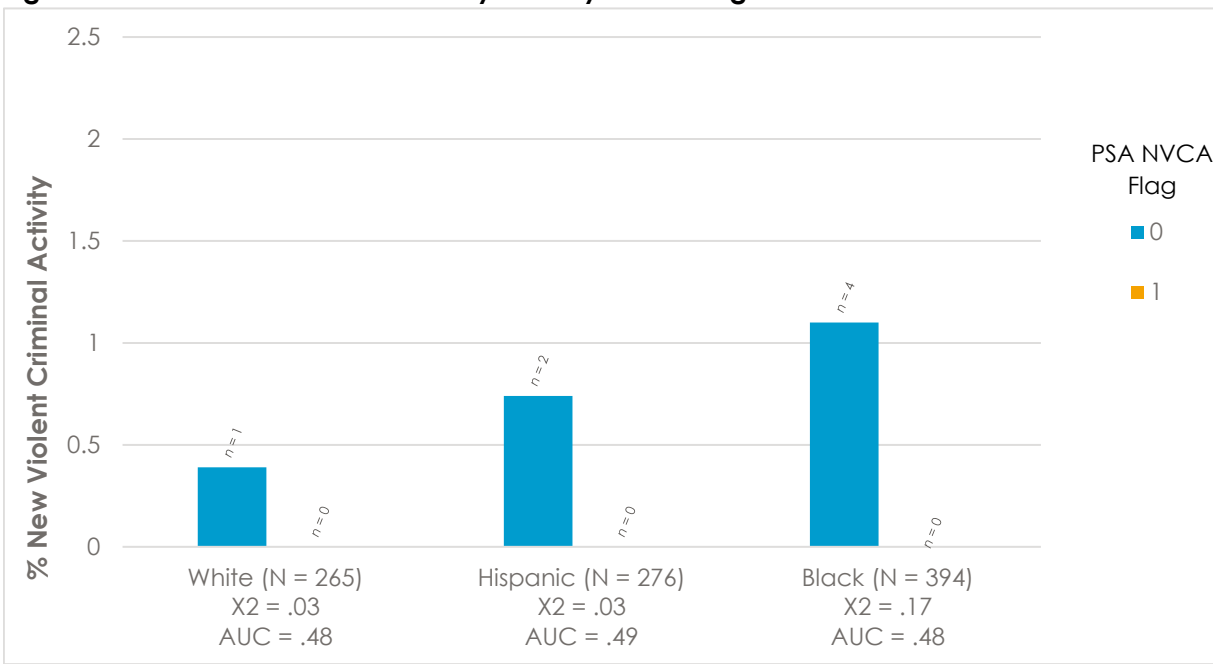
Figure 7 presents NVCA rates by PSA NVCA flag and race/ethnicity based on the sample of 990 defendants from Contra Costa County. This analysis provides insight into how NVCA rates differ across racial/ethnic subgroups in relation to the PSA NVCA flag. Of note, no defendants of an "other" race or identity committed a new violent offense and are therefore not included in this analysis.

For White defendants with an NVCA flag of 0 (i.e., no flag), the rate of new violent criminal activity is 0.4%. No White defendants with an NVCA flag of 1 committed a new violent offense. The Chi-square test ($X^2 = 0.03$, $p > .05$) does not indicate a statistically significant association between the NVCA flag and violent criminal activity for White defendants, and the AUC of .48 suggests poor predictive accuracy for the PSA NVCA flag in this group.

Among Hispanic defendants (N = 276), a similar trend is observed. The NVCA rate for those with an NVCA flag of 0 is 0.7%, and no Hispanic defendants with an NVCA flag of 1 committed a new violent offense. The Chi-square test ($X^2 = 0.03$, $p > .05$) does not show a statistically significant relationship between the PSA NVCA flag and NVCA outcomes, and the AUC of .49 indicates limited predictive power for the PSA flag among Hispanic defendants.

For Black defendants (N = 394), the NVCA rate is slightly higher than other subgroups. Among Black defendants with an NVCA flag of 0, the rate of new violent criminal activity is 1%, while no Black defendants with an NVCA flag of 1 committed a new violent offense. The Chi-square test ($X^2 = .17$, $p > .05$) does not show a statistically significant relationship between the NVCA flag and NVCA outcomes for Black defendants, and the AUC of .48 suggests poor predictive accuracy in this group.

Figure 7. New Violent Criminal Activity Rate by NVCA Flag & Race[a]



[a] No defendants categorized as an "Other" racial or ethnic identity failed.

* p < .05, ** p < .01, *** p < .001
n values reflect the number of defendants who failed in each category.

In summary, across all racial and ethnic subgroups examined, the PSA NVCA flag does not show a meaningful association with new violent criminal activity outcomes, as evidenced by the lack of statistical significance in Chi-square tests and low AUC scores (close to .5) across groups. This suggests the limited utility of the PSA NVCA flag in predicting violent offense risk within each racial/ethnic subgroup.

To further investigate any bias in the tool's accuracy, the evaluation team again incorporated methods utilized in the California Policy Lab's San Francisco County PSA validation.[22] Specifically, the RDA team replicated the analysis previously displayed in **Table 11**, examining accuracy, false negatives, and false positives for each racial or ethnic group.[23] Once again, scaled scores greater than three are predicted failures. In contrast, scaled scores less than three are predicted non-failures. **Tables 13 – 15** display the accuracy, false negative, and false positive rates for the sample overall and each racial or ethnic group. The highest rate in each category is identified with orange fill, while the lowest is blue.

When interpreting these correct classification results by demographic sub-groups, the evaluation team was particularly concerned with identifying instances in which one racial or ethnic group was classified as high risk at a greater relative rate while also having a greater relative false alarm rate. Observing this dynamic would indicate that the risk for that racial or ethnic group is overstated, potentially indicating tool bias. Results from this analysis did not detect any significant bias concerns, conforming with the findings in the proceeding paragraphs. Results again are somewhat limited by small sample sizes, discussed in this report's limitations section.

Across PSA predicted outcomes, defendants with an "other" racial or ethnic identity are categorized as high-risk at the lowest rate. Consistently, defendants of an "other" identity have also been classified with the highest accuracy rate and have the lowest false negative rate, i.e., missed failures.[24] Despite the low rate of defendants with an "other" identity assessed as high-risk (29%), these defendants have the highest false positive rate (i.e., false alarm for FTAs at 57%), indicating an even smaller share of defendants should have been identified as high-risk.

White defendants are categorized as high-risk at the highest rate for FTA, and NCA predicted outcomes, while Black defendants are categorized as high-risk at the highest rate for NVCA. However, when White and Black defendants are categorized as high-risk at the highest rate, they have been classified with the lowest accuracy and the highest false negative rate (i.e., missed failures), indicating an even greater share of defendants should have been identified as high-risk based on observed results in the validation sample.

[22] Skog, A., Lacoe, J. (2021). Validation of the PSA in SF.
[23] Refer to **Appendix H** for the complete replication results associated with Tables 13 – 15.
[24] The accuracy rate is calculated as the sum of true failures and non-failures divided by the total population of failures and non-failures. The false negative rate is calculated by dividing false non-failures by all failures; it represents the share of failures that were not predicted (i.e., missed fail rate).

Table 13. FTA Classification Rates, by Race

	N	High-Risk	%	Accuracy (Correct Classification True Failures & Non-Failures)	False Negative Rate (Missed Fail)	False Positive Rate (False Alarm)
Overall	**990**	**406**	**41%**	**59%**	**38%**	**50%**
White	265	141	53%	51%	52%	44%
Black or African American	394	147	37%	61%	34%	54%
Hispanic	276	102	37%	63%	32%	51%
"Other" Identity	55	16	29%	67%	24%	57%

Note: Cells with the highest rate across racial groups are filled with orange, and cells with the lowest rate are filled with blue; Defendants classified as "high risk" for FTA had an FTA scaled score of at least 3.

Table 14. NCA Classification Rates, by Race

	N	High-Risk	%	Accuracy (Correct Classification True Failures & Non-Failures)	False Negative Rate (Miss Future Fail)	False Positive Rate (False Alarm)
Overall	**990**	**464**	**47%**	**54%**	**46%**	**46%**
White	265	149	56%	43%	57%	53%
Black or African American	394	195	49%	53%	48%	39%
Hispanic	276	102	37%	63%	35%	50%
"Other" Identity	55	18	33%	67%	31%	50%

Note: Cells with the highest rate across racial groups are filled with orange, and cells with the lowest rate are filled with blue; Defendants classified as "high risk" for NCA had an NCA scaled score of at least 3.

Table 15. NVCA Classification Rates, by Race

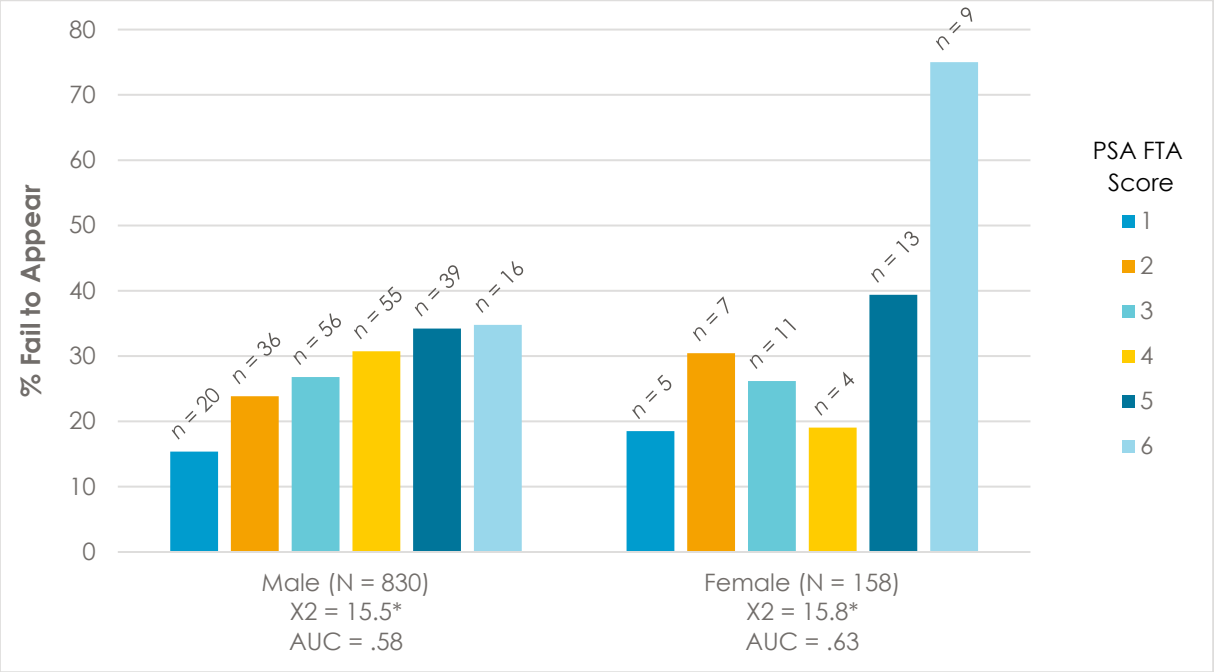
	N	High-Risk	%	Accuracy (Correct Classification True Failures & Non-Failures	False Negative Rate (Miss Future Fail)	False Positive Rate (False Alarm)
Overall	**990**	**30**	**3%**	**96.3%**	**3.1%**	**100.0%**
White	265	10	4%	95.8%	3.8%	100.0%
Black or African American	394	16	4%	94.9%	4.1%	100.0%
Hispanic	276	4	1%	97.8%	1.5%	100.0%
"Other" Identity	55	0	0%	100.0%	0.0%	[Omitted]

Note: Cells with the highest rate across racial groups are filled with orange, and cells with the lowest rate are filled with blue; Defendants classified as "high risk" for NVCA had an NVCA flag of "yes."

Predictive Validity by Gender

Figure 8 illustrates FTA rates by PSA scaled score and gender for the sample of interest. For male defendants (*N* = 830), the FTA rates increase steadily as PSA scaled scores rise. The FTA rate for males with a PSA scaled score of 1 is 15% (n = 20), and it gradually climbs as the PSA scaled score increases. Male defendants with a PSA scaled score of 2 have an FTA rate of 24% (n = 36), which

Figure 8. Failure to Appear Rate by PSA Scaled Score & Gender



* p < .05, ** p < .01, *** p < .001

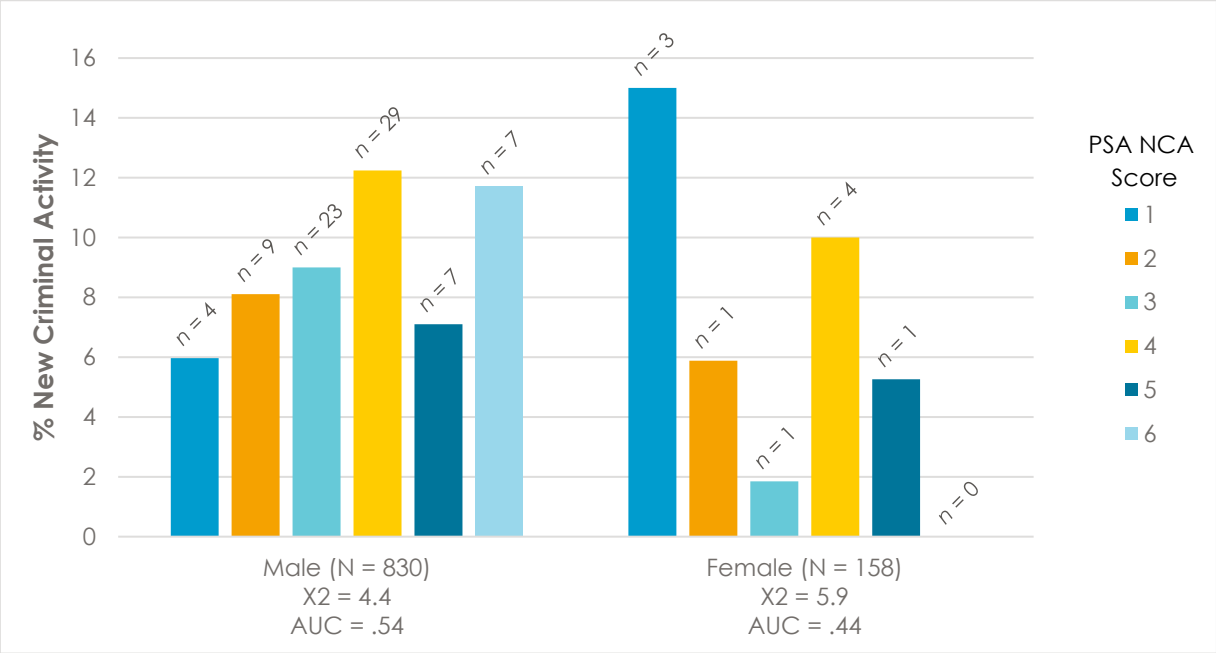
n values reflect the number of defendants who failed in each category.

rises to 27% (n = 56) for a PSA scaled score of 3. The FTA rate reaches 34% (n = 39) for a PSA scaled score of 5, and individuals with a PSA scaled score of 6 exhibit an FTA rate of 35% (n = 16). The Chi-square test ($X^2 = 15.4$, $p < .01$) shows a statistically significant association between PSA scaled score and FTA for male defendants, with an AUC of 0.58, indicating fair predictive accuracy.

For female defendants ($N = 158$), the FTA rate increases somewhat irregularly with higher PSA scaled scores but exhibits a more dramatic rise at higher PSA scaled scores. Females with a PSA scaled score of 1 have an FTA rate of 19% (n = 5), while those with a PSA scaled score of 3 have a rate of 26% (n = 11). The FTA rate jumps significantly for females with PSA scaled scores of 5 and 6, with FTA rates of 39% (n = 13) and 75% (n = 9), respectively. The Chi-square test ($X^2 = 15.8$, $p < .01$) shows a significant relationship between PSA scaled score and FTA for females, with a higher AUC of 0.63, indicating fair to good predictive accuracy, better than male defendants.

In summary, both male and female defendants exhibit higher FTA rates as PSA scaled scores increase, confirming the PSA scaled score as a valid predictor of failure to appear. However, the relationship between PSA scaled score and FTA rate appears stronger for females, particularly at higher PSA scaled scores, as reflected in the steeper rise in FTA rates and the higher AUC value.

Figure 9. New Criminal Activity by PSA Scaled Score & Gender



* p < .05, ** p < .01, *** p < .001
n values reflect the number of defendants who failed in each category.

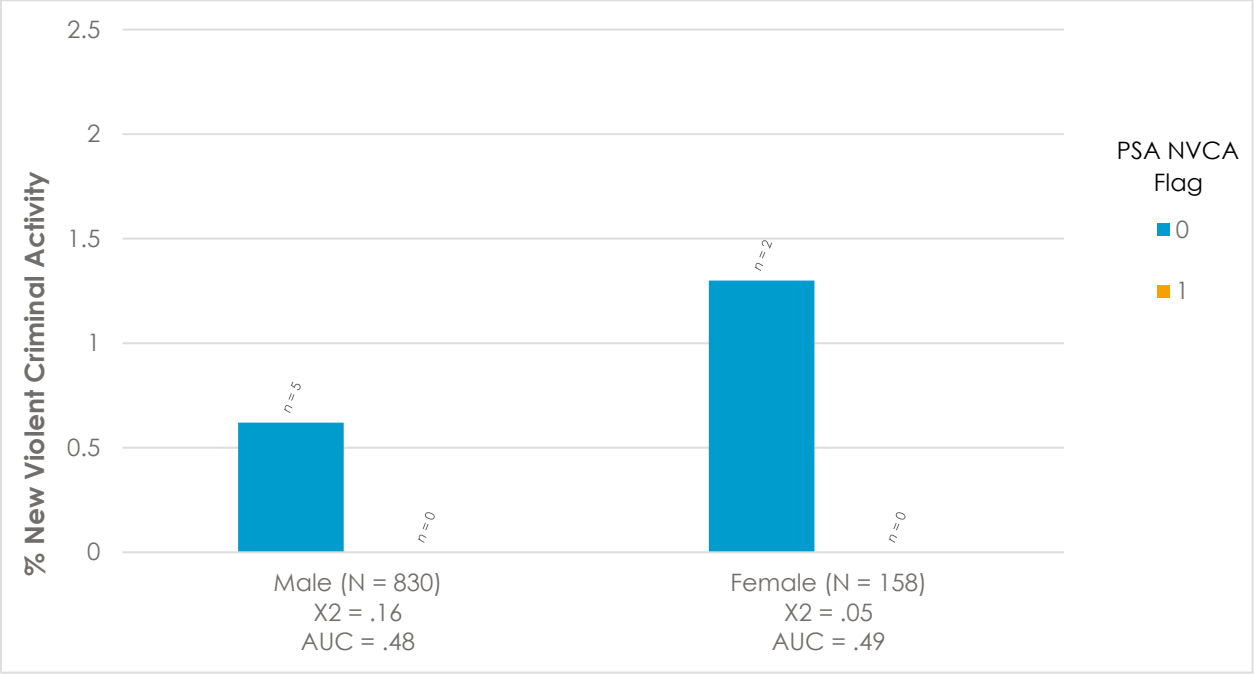
This suggests that while PSA scaled scores help predict FTA for both genders, they may be slightly more accurate for females in this sample.

Figure 9 displays the NCA rate by PSA scaled score and gender for the sample of 983 defendants from Contra Costa County. For male defendants (*N* = 825), the NCA rates vary across PSA scaled scores, generally increasing with higher scores. The NCA rate for males with a PSA scaled score of 1 is 6% (*n* = 4), while for those with a PSA scaled score of 2, the rate increases to 8% (*n* = 9). The highest rates are observed at PSA scaled scores of 3 and 4, with NCA rates of 9% (n = 23) and 12% (n = 29), respectively. The rate decreases slightly for PSA scaled scores of 5 (7%, *n* = 7) and 6 (12%, *n* = 7). The Chi-square test ($X^2 = 4.4$, $p > .05$) indicates no significant relationship between PSA scaled scores and NCA for males, and the AUC of 0.54 suggests less than fair, or slight, predictive accuracy for this group.

The NCA rates follow a different pattern for female defendants (*N* = 158). The NCA rate is 15% (*n* = 3) for females with a PSA scaled score of 1, decreasing irregularly to 0% for a scaled score of 6. The Chi-square test ($X^2 = 5.9$, $p > .05$) also shows no significant relationship between PSA scaled scores and NCA for females, and the AUC of 0.44 indicates low predictive accuracy for the PSA scaled score in predicting new criminal activity among female individuals.

In summary, while the PSA scaled score shows a slight relationship with new criminal activity for male defendants, with higher PSA scaled scores corresponding to higher NCA rates, the same trend is not apparent for female individuals. The PSA scaled score appears to have limited predictive utility for forecasting new criminal activity, especially for females, as reflected in the lower AUC score. Overall, the predictive accuracy of PSA scaled scores for NCA is relatively weak for both genders in this sample, with a slightly stronger association observed among male defendants.

Figure 10. New Violent Criminal Activity Rate by NVCA Flag & Gender



* p < .05, ** p < .01, *** p < .001
n values reflect the number of defendants who failed in each category.

Figure 10 presents the rate of NVCAs by PSA NVCA flag and gender for the sample of interest. For males (N = 830), the NVCA rate is low, with 5 out of 830 defendants flagged as low risk (i.e., NVCA flag value of 0) committing a new violent offense. This results in an NVCA rate of 0.62% among low-risk males. No males flagged as high risk (i.e., NVCA flag value of 1) committed a new violent offense. The Chi-square test ($X^2 = 1.6$, $p > .05$) shows no significant association between the NVCA flag and actual NVCA outcomes for males, and the AUC of .48 suggests minimal predictive power for the PSA NVCA flag and this subgroup.

Among females (N = 158), the NVCA rate was slightly higher, with 2 out of 158 defendants flagged as low risk of committing a new violent offense, resulting in an NVCA rate of 1.3%. Like the male subgroup, no females flagged as high risk committed a new violent offense. The Chi-square test ($X^2 = .05$, $p > .05$) indicates no significant relationship between the NVCA flag and NVCA outcomes for females, and the AUC of .49 reflects poor predictive accuracy.

In summary, across both male and female subgroups, the PSA NVCA flag does not effectively differentiate between individuals who are likely and unlikely to commit new violent offenses. For both genders, none of the individuals flagged as high risk committed a new violent offense, while

a small proportion of those flagged as low risk did. The lack of significant Chi-square values and AUC scores close to 0.5 suggest that the PSA NVCA flag has limited utility in predicting violent criminal activity for either gender.

III. Summary and Recommendations

PSA results are just one factor the Probation Department considers when making their pretrial release recommendation to the Court. The Probation Department and the Judicial Officer of the Court (i.e., the judge) exercise their professional discretion when interpreting the PSA results, with the judge having the ultimate authority to release defendants pretrial.

Released defendants were very successful in Contra Costa County. Looking at the total validation sample used for this study, **most released defendants in the predictive validity sample did not have an FTA (73%, n = 719), NCA (91%, n = 901), or NVCA (99%, n = 983). However, this study does find that the PSA has fair to weak predictive accuracy. The RDA evaluation team did not detect serious bias concerns by different defendant identities.** The PSA's FTA scaled score has fair predictive accuracy for the total validation sample and all demographic groups, reflecting the high confidence survey respondents have in the tool's ability to predict FTAs accurately. The PSA has weak predictive accuracy for NCA and NVCA across the full sample. Similarly, NCA scaled scores and the NVCA flag were weak predictors of their respective outcomes across different defendant identities, mirroring PSA's limited predictive accuracy for those outcomes across the full sample as opposed to identifying bias concerns.

Table 16 shows the AUC results from prior PSA validation studies in the region and statewide to contextualize these results. For reference, AUC minimum value cut-offs used in this study are as follows: fair (AUC-ROC = .55), good (AUC-ROC = .64), and excellent (AUC-ROC = .71). AUCs for the California Policy Lab's San Francisco County study fell within the fair to good range for predictive accuracy with a sample size of 9,881 defendants.[25] The Judicial Council's statewide study AUCs fell within the fair to excellent range for predictive accuracy with a sample size of 37,949 defendants.[26] In contrast, this Contra Costa County sample study size (i.e., 990 defendants) is one tenth the San Francisco County sample size and less than 3% of the statewide sample size. Due to San Francisco's proximity to Contra Costa County, it is reasonable to expect the PSA to behave at least similarly to one another. Given the limitations presented in this validation study, San Francisco's previous AUCs may paint a more accurate picture of the PSA's predictive validity.

Table 16. Local Comparison of Validation Results, AUC-ROC values

Predicted Outcome	Contra Costa County, 2020-2023	San Francisco County, 2016-2019 [1]	California, 2019-2021 [2]
FTA	.59	.63	.68
NCA	.53	.62	.71
NVCA	.48	.66	.58

Note. PSA = Public Safety Assessment; AUC-ROC = area under the receiver operating characteristic curve. (1) Skog, A., Lacoe, J. (2021). Validation of the PSA in San Francisco. *California Policy Lab.* (2) Judicial Council. (2022). Pretrial risk assessment tool validation: Pretrial pilot program. *Judicial Council of CA, Operations and Program Division, Criminal Justice Services.*

At the local level, the AUCs for the FTA predicted outcome is the most similar across the two counties, representing fair accuracy. The most substantial differences exist between the NVCA

[25] AUC values in San Francisco did not reach the 0.71 minimum threshold for excellent predictive validity for any PSA outcomes. Skog, A., Lacoe, J. (2021). Validation of the PSA in SF.

[26] AUC values across the state were calculated using data from the following counties: Los Angeles (n = 24,982), Sacramento (n = 10,132), Sonoma (n = 1,546), Tulare (n = 1,146), Calaveras (n = 115), and Tuolumne (n = 28). Judicial Council. (2022). Pretrial risk assessment tool validation.

AUCs. The Contra Costa County NVCA of 0.48 is weaker predictive accuracy than a coin toss, while San Francisco's AUC indicates moderate predictive accuracy. Of note, the NVCA AUC was also the weakest of the three predicted outcomes in the statewide study.

Recommendations

Ultimately, the weaker predictive accuracy identified in this study may stem from several factors mentioned in the **Limitations** section. As a result, the findings should be carefully interpreted. Additionally, the evaluation team cannot offer scoring changes to the PSA based on these results alone. In partnership with the Pretrial Services Unit, RDA has developed a list of recommendations that will address the findings presented in this report.

- Continue efforts already underway, such as data tracking systems to alert officers to complete data entry, detailed judicial agreement collection, and incorporation of Court data that will facilitate consistent outcome reporting for all assessed defendants. Additionally, RDA recommends enhancing accurate outcome tracking with data automations and data validation processes that will allow Probation to continuously assess data quality.
- Ensure consistency in item scoring based on identified areas of concern. This can include booster training sessions with officers in the unit to provide a clear understanding of which offenses are considered violent for PSA scoring and systematic outcome tracking. To implement this training plan, RDA specifically recommends Probation prepares a January training to review proper scoring for risk factors with identified issues (e.g., items assessing current or previous violent offenses). Probation should plan for a second training in March to review case management system updates and automations to ensure consistent and efficient data entry. A third training should be planned within six months as a group exercise comparing individual officer PSA scoring for three vignettes (i.e., fictional defendants and cases) against a trainer and unit supervisor scoring. Ultimately, these initial trainings are geared towards improving officer scoring accuracy and consistency to boost the PSA's local predictive validity. Continuing with regular boosters after these initial training sessions will further promote scoring accuracy and consistency.
- Consult with Advancing Pretrial Policy and Research (APPR), a PSA technical assistance provider contracted by Arnold Ventures, to coordinate unit trainings, discuss the issues identified with individual predictive accuracy of risk factors (e.g., risk factors three and seven), and any improvements that can be made (e.g., weight changes) to improve the PSA's accuracy.
- Conduct and report validation results to stakeholders again in three years to continue Probation's transparency efforts, which can help in maintaining trust and accountability while remaining compliant with SB 36 requirements. With efforts underway to improve data collection and officer scoring, RDA recommends Probation perform an initial data study in two years (i.e., 2027) with the goal of determining whether there are sufficient observations to conduct another validation (e.g., at least 1,000 observations with at least 200 observations for each demographic sub-group of interest). Given the weaker predictive accuracy results identified in this study, it would be preferrable to conduct the next validation with data collected over the two-year period from 2025 to 2026 if sample sizes are large enough. If sample sizes are not yet large enough within two years, the next validation will take place in 2028 with data drawn from 2025-2027.

Acknowledgments

RDA would like to thank Contra County Probation and their points of contact, Bart Sloan, Chad Cardoza, Greg Kurlinski, and Yuri Secoquian, for facilitating this project and offering their valuable thought partnership. We would also like to thank the officers who participated in the survey for their responses. This report was written by Project Manager Penelope Fergison, M.P.P., Analyst Jennifer Lux, Ph.D., and Project Sponsor Carrie Coen, M.A.

Appendix A: Contra Costa County PSA Staff Survey

PSA Administration & Use of Results[27]

1) How often do you administer the Public Safety Assessment (PSA)?*

( ) 5 or more times a week

( ) 1-4 times a week

( ) 2-3 times a month

( ) Once per month

( ) Less than once per month

( ) Other - Write In (Required): _____*

2) Do you consult data available in existing court and probation computer systems when scoring the PSA (e.g., CLETS)? If yes, please provide examples of databases or records consulted.*

( ) Yes

( ) No

( ) Other - Write In (Required): _____*

3) Do you reference scoring instructions when scoring a PSA assessment? If not, please provide an explanation.*

( ) Yes

( ) No

( ) Other - Write In (Required): _____*

4) Do you reference the release conditions matrix (see below) when determining a defendant's release level from FTA and NCA scaled scores? If not, please provide an explanation.*

( ) Yes - Write In (Required): _____*

( ) No

( ) Other - Write In (Required): _____*

[27] An asterisk (*) indicates a question is required.

	Public Safety Assessment Release Conditions Matrix						
Failure to Appear (FTA) Scaled Score ↓	New Criminal Activity (NCA) Scaled Score →	1	2	3	4	5	6
1		1	1				
2		1	1	1	2	3	
3			2	2	2	3	4
4			2	3	3	4	4
5			2	3	4	4	5
6					4	5	5

5) Do you reference the defendant's release level (see below) to assist in release decision recommendations for the pretrial report? If not, please provide an explanation.*

( ) Yes

( ) No

( ) Other - Write In (Required): _____*

			Pretrial Release Level			
Release Activities and Conditions	1	2	3	4	5 NVCA-No Flag	5 NVCA- With Flag
Mandatory Statutory Conditions	X	X	X	X	X	
Court Reminders	X	X	X	X	X	
Criminal History Checks Once Per Month			X	X	X	
Check-In Once per Month			X			
Check-in Twice per Month				X	X	
Call in Once per Month			X			
Other Case-Specific Conditions	X	X	X	X	X	
Electronic Home Detention (EHD)				As Ordered by the Court	X	
SCRAM (DUI)	As Ordered by the Court	As Ordered by the Court	As Ordered by the Court	As Ordered by the Court	As Ordered by the Court	
NO PTS Release						X

6) How often does the judge agree with your recommended release decision (i.e., to release or detain a defendant pretrial)?*

_____

7) What are the common reasons the judge disagrees with your recommended release decision?*

_____

8) In cases where pretrial release is granted, how often does the judge adopt your recommended release terms in the pretrial report (e.g., check-in frequency, substance use)?*

_____

9) What are the common reasons the judge modifies your recommended release terms?*

_____

Quality Assurance & Release Decision Framework

10) Have you received formal training on how to use the PSA?*

( ) Yes

( ) No

( ) Other - Write In (Required): _____*

11) Have you received any retraining on the PSA (i.e., attended a recertification or booster session)?*

( ) Yes

( ) No

( ) Other - Write In (Required): _____*

12) If yes, how often do these trainings occur?*

( ) Once every six months

( ) Annually

( ) Once every two years

( ) Other - Write In (Required): _____*

13) Is there a quality assurance process (e.g., supervisor approval or having someone observe your interview and scoring) in place concerning the PSA?*

( ) Yes

( ) No

( ) Other - Write In (Required): _____*

14) In addition to the PSA results, what other pieces of information do you consider as you're making your release recommendations in the pretrial report (e.g., instant offense)?*

_____

15) How do you prioritize these pieces of information (e.g., PSA, instant offense, victim statement, public safety) when developing a release recommendation (i.e., what carries more weight)?*

_____

PSA Fairness, Benefits, & Staff Support

16) Please indicate your level of agreement with the following statements related to the PSA as it is used by your agency.*

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Uncertain
The PSA has helped our agency establish a common language regarding defendants' risk level	( )	( )	( )	( )	( )	( )
The PSA is easy to read, interpret, and use	( )	( )	( )	( )	( )	( )
The PSA increases consistency in rules, policies, and procedures regarding assessment practices	( )	( )	( )	( )	( )	( )

The PSA reduces variation in decision-making across individual staff	( )	( )	( )	( )	( )	( )
The PSA helps our agency make proper decisions regarding whether a defendant is appropriate for pretrial supervision versus custody	( )	( )	( )	( )	( )	( )
The PSA reduces bias in criminal justice decision making in our agency	( )	( )	( )	( )	( )	( )

17) Please indicate your level of agreement with the following statements related to *usefulness* of the PSA at your agency.*

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Uncertain
The PSA has made my job easier	( )	( )	( )	( )	( )	( )
There is staff support for implementing the PSA	( )	( )	( )	( )	( )	( )
There are clear guidelines for when to use the PSA	( )	( )	( )	( )	( )	( )
There is a protocol for how to use the PSA	( )	( )	( )	( )	( )	( )

18) How confident are you that the results produced by the PSA accurately captures the likelihood of the following:*

	Completely lack confidence	Not confident	Neutral	Confident	Completely confident	Uncertain
Defendant missing a pretrial court hearing	( )	( )	( )	( )	( )	( )
Defendant being arrested for a new criminal offense while released	( )	( )	( )	( )	( )	( )
Defendant being arrested for a new violent offense while released	( )	( )	( )	( )	( )	( )

19) How confident are you that a defendant will receive the same scores no matter which staff member conducted the assessment.*

( ) Completely lack confidence

( ) Not confident

( ) Neutral

( ) Confident

( ) Completely confident

( ) Uncertain - Write In (Required): _____*

20) How do you rate your overall satisfaction with the PSA?*

( ) Very Dissatisfied

( ) Dissatisfied

( ) Neutral

( ) Satisfied

( ) Very Satisfied

( ) Other - Write In (Required): _____*

PSA Strengths & Challenges

21) List 3 challenges of using the PSA in your agency. This question is required, please write "NA" when you cannot think of a challenge.*

Challenge 1: _____

Challenge 2: _____

Challenge 3: _____

22) List 3 strengths of using the PSA in your agency. This question is required, please write "NA" when you cannot think of a strength.*

Strength 1: _____

Strength 2: _____

Strength 3: _____

Closing Thoughts

23) Is there anything else you would like to share with us about the PSA or its use in the department?

_____

Staff Information

24) What is your highest level of education completed?*

( ) Bachelor's degree

( ) Graduate degree

25) How many years of experience do you have working in the criminal justice field (including your time at probation)?*

_____

26) How long have you been working at probation?*

_____

27) How long have you been working in your current role in the Pretrial Unit?*

_____

28) How many defendants are currently on your caseload?*

_____

Thank You!

Appendix B: Detailed Defendant Characteristics, by Pretrial Release Decision

	Overall (*N* = 2,054)	Released (*N* = 1,114, 54%)		Denied & Other (*N* = 940, 46%)	
	n	n	%	n	%
Gender					
Male	1,697	930	54.8%	767	45.2%
Female	353	182	51.6	171	48.4
Race					
White	566	293	51.8%	273	48.2%
Hispanic/Latino	551	304	55.2	247	44.8
Black/African American	818	457	55.9	361	44.1
API/American Indian/'Other' Identity	119	60	50.4	59	49.6
FTA Scaled Score **					
1	289	180	62.3%	109	37.7%
2	382	194	50.8	188	49.2
3	498	283	56.8	215	43.2
4	434	233	53.7	201	46.3
5	310	162	52.3	148	47.7
6	141	62	44.0	79	56.0
NCA Scaled Score ***					
1	169	102	60.4%	67	39.6%
2	231	143	61.9	88	38.1
3	625	354	56.6	271	43.4
4	572	305	53.3	267	46.7
5	286	135	47.2	151	52.8
6	171	75	43.9	96	56.1
NVCA Flag **					
0 (i.e., no flag)	1,970	1,082	54.9%	888	45.1%
1 (i.e., NVCA flag)	84	32	38.1	52	61.9

* p < .05, ** p < .01, *** p < .001

Appendix C: Detailed Defendant Characteristics, by Judicial Officer Agreement with Probation Recommendations

	Overall	Agreement (*N* = 867, 47%)		Disagreement (*N* = 993, 53%)	
	n	n	%	n	%
Gender					
Male	1,539	719	47%	820	53%
Female	317	146	46%	171	54%
Race					
White	515	228	44%	287	56%
Hispanic/Latino	508	240	47%	268	53%
Black/African American	728	351	48%	377	52%
API/American Indian/'Other' Identity	109	48	44%	61	56%
FTA Scaled Score ***					
1	250	133	53%	117	47%
2	326	151	46%	175	54%
3	446	219	49%	227	51%
4	413	208	50%	205	50%
5	292	103	35%	189	65%
6	133	53	40%	80	60%
NCA Scaled Score					
1	146	75	51%	71	49%
2	200	99	50%	101	51%
3	558	273	49%	285	51%
4	525	235	45%	290	55%
5	273	114	42%	159	58%
6	158	71	45%	87	55%
NVCA Flag					
0 (i.e., no flag)	1,787	835	47%	952	53%
1 (i.e., NVCA flag)	73	32	44%	41	56%

* p < .05, ** p < .01, *** p < .001

Appendix D: Logistic Regression Results, by Outcome

Table E1. Binomial Logit Model PSA Scaled Score & Failure to Appear

	FTA	FTA (ME)
FTA scaled score	.22***	.04***
	(.05)	(.01)
Days between release & close	-.00**	
	(.00)	
Constant	-1.31***	
	(.23)	
Observations	814	814

Note. ME = marginal effects.
Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Table E2. Binomial Logit Model PSA scaled Score & New criminal activity

	NCA	NCA (ME)
NCA scaled score	.03	.00
	(.09)	(.01)
Days between release & close	-.00*	
	(.00)	
Constant	-1.96***	
	(.36)	
Observations	814	814

Note. ME = marginal effects.
Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Table E3. Binomial Logit Model NVCA Flag & NVCA Outcome

	NVCA	NVCA (ME)
NVCA flag	.00	.00
	(.)	(.)
Days b/w Release & Close	-.01	
	(.01)	
Constant	-3.35***	
	(.55)	
Observations	794	794

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Appendix E: Combined Factor Analysis Results, by Individual Risk Factor

	Risk Factor		Score	N	FTA Rate	Bivariate (Log Odds Ratio)	Multivariate (Log Odds Ratio)	N	NCA Rate	Bivariate (Log Odds Ratio)	Multivariate (Log Odds Ratio)	N	NVCA Rate	Bivariate (Log Odds Ratio)	Multivariate (Log Odds Ratio)
Item 1	Age at current arrest	23 or older	0					840	9.4%	**-0.13**	0.02				
		22 or younger	2					150	7.3%						
Item 2	Current violent offense	No	0									929	**0.8%**	[Omitted]	[Omitted]
		Yes	2									61	**0.0%**		
Item 2A	Current violent offense and 20 years old or younger	No	0									984	**0.7%**	[Omitted]	[Omitted]
		Yes	1									6	**0.0%**		
Item 3	Pending charge at the time of the arrest	No	0	769	26.3%	0.24	0.02	769	**9.1%**	**-0.02**	**-0.04**	769	**0.8%**	**-0.55**	**-0.60**
		Yes	1^	221	31.2%			221	**8.6%**			221	**0.5%**		
Item 4	Prior misdemeanor conviction	No	0					336	6.8%	0.42	0.27				
		Yes	1					654	10.1%						
Item 5	Prior felony conviction	No	0					474	7.8%	0.28	0.12				
		Yes	1					516	10.1%						
Item 5A	Prior conviction (misdemeanor or felony)	No	0	261	20.7%	0.49**	0.28					261	**0.8%**	**-0.11**	0.24
		Yes	1	729	29.8%							729	**0.7%**		
Item 6	Prior violent conviction	No	0					782	**8.8%**	0.04	**-0.11**	782	**0.9%**	[Omitted]	[Omitted]
		Yes, 1-2	1					190	**10.0%**			190	**0.0%**		
		Yes, 3+	2					18	**5.6%**			18	**0.0%**		
Item 7	Prior failure to appear in the past 2 years	No	0	587	23.2%	0.19***	0.04***	587	**9.0%**	0.04	0.02				
		Yes, 1	2^	217	28.6%			217	**7.8%**						
		Yes, 2+	4^	185	39.5%			185	**10.3%**						
Item 8	Prior failure to appear older than 2 years	No	0	495	23.2%	0.42**	0.08**								
		Yes	1	495	31.5%										
Item 9	Prior sentence to incarceration	No	0					298	6.7%	0.22	0.11				
		Yes	2					690	10.0%						

Note: ^ Item 3 score is 0 for "No" responses across all three outcomes while a "Yes" response is a 1 for FTA and NVCA and a 3 for NCA. Item 7 score is 0 for "No" responses across FTA and NCA while a "Yes, 1" is a 1 for NCA and "Yes, 2+" is a 2 for NCA.

Appendix F: Individual Risk Factor Analysis Logistic Regression Results, by Outcome & Defendant Characteristics

Table F1. Univariate & Multivariate Binomial Logit Model Factor Scores & FTA Failure

	Univariate	Univariate (ME)	Multivariate, Joint Test	Multivariate Joint Test (ME)
Item 3	0.24 (0.17)	0.05 (0.03)	0.02 (0.18)	0.00 (0.03)
Item 5a	0.49** (0.17)	0.10** (0.03)	0.28 (0.20)	0.05 (0.04)
Item 7	0.19*** (0.04)	0.04*** (0.01)	0.16*** (0.05)	0.03*** (0.01)
Item 8	0.42** (0.14)	0.08** (0.03)	0.20 (0.17)	0.04 (0.03)
Constant			-1.50*** (0.16)	
Observations	990	990	989	989

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table F2. Univariate & Multivariate Binomial Logit Model Factor Scores & NCA Failure

	Univariate	Univariate (ME)	Multivariate, Joint Test	Multivariate Joint Test (ME)
Item 1	-0.13 (0.17)	-0.01 (0.01)	0.02 (0.19)	0.00 (0.02)
Item 3	-0.02 (0.09)	-0.00 (0.01)	-0.04 (0.10)	-0.00 (0.01)
Item 4	0.42 (0.25)	0.03 (0.02)	0.27 (0.37)	0.02 (0.03)
Item 5	0.28 (0.23)	0.02 (0.02)	0.12 (0.28)	0.01 (0.02)
Item 6	0.04 (0.24)	0.00 (0.02)	-0.11 (0.26)	-0.01 (0.02)
Item 7	0.04 (0.14)	0.00 (0.01)	0.02 (0.15)	0.00 (0.01)
Item 9	0.22 (0.13)	0.02 (0.01)	0.11 (0.21)	0.01 (0.02)
Constant			-2.68*** (0.30)	
Observations	988-990	988-990	987	987

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table F3. Univariate & Multivariate Binomial Logit Model Factor Scores & NVCA Failure

	Univariate	Univariate (ME)	Multivariate, Joint Test	Multivariate Joint Test (ME)
Item 2	0.00 (.)	0.00 (.)	0.00 (.)	0.00 (.)
Item 2a	0.00 (.)	0.00 (.)	0.00 (.)	0.00 (.)
Item 3	-0.55 (1.08)	-0.00 (0.01)	-0.60 (1.09)	-0.01 (0.01)
Item 5a	-0.11 (0.84)	-0.00 (0.01)	0.24 (0.84)	0.00 (0.01)
Item 6	0.00 (.)	0.00 (.)	0.00 (.)	0.00 (.)
Constant			-4.70*** (0.72)	
Observations	782-929	782-929	736	736

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix G: Logistic Regression Results, by Outcome & Defendant Characteristics

Table G1. Binomial Logit Model PSA Scaled Score & FTA by Race

	FTA	FTA (ME)
FTA scaled score	.22***	
	(.06)	
Days b/w Release & Close	-.00**	
	(.00)	
Race: Hispanic	.01	
	(.21)	
Race: Black	-.24	
	(.20)	
Race: Other	.02	
	(.37)	
Race: White (ME)		.04***
		(.01)
Race: Hispanic (ME)		.04***
		(.01)
Race: Black (ME)		.04***
		(.01)
Race: Other (ME)		.04***
		(.01)
Constant	-1.22***	
	(.27)	
Observations	814	814

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Table G2. Binomial Logit Model PSA Scaled Score & NCA by Race

	NCA	NCA (ME)
NCA scaled score	.06	
	(.09)	
Days b/w Release & Close	-.00*	
	(.00)	
Race: Hispanic	.73*	
	(.34)	
Race: Black	.56	
	(.32)	
Race: Other	.62	
	(.55)	
Race: White (ME)		.00
		(.01)
Race: Hispanic (ME)		.01
		(.01)
Race: Black (ME)		.01
		(.01)
Race: Other (ME)		.01
		(.01)
Constant	-2.54***	
	(.46)	
Observations	814	814

Table G3. Binomial Logit Model NVCA Flag & NVCA Outcome by Race

	NVCA	NVCA (ME)
NVCA flag	.00	
	(.)	
Days b/w Release & Close	-.01	
	(.01)	
Race: Hispanic	.50	
	(1.24)	
Race: Black	.98	
	(1.13)	
Race: Other	.00	
	(.)	
Race: White (ME)		.00
		(.)
Race: Hispanic (ME)		.00
		(.)
Race: Black (ME)		.00
		(.)
Race: Other (ME)		.00
		(.)
Constant	-3.95***	
	(1.10)	
Observations	750	750

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Table G4. Binomial Logit Model PSA Scaled Score & Failure by Gender

	FTA	FTA (ME)
FTA scaled score	.22***	
	(.05)	
Days b/w Release & Close	-.00**	
	(.00)	
Female	.18	
	(.21)	
Sex: Male (ME)		.04***
		(.01)
Sex: Female (ME)		.05***
		(.01)
Constant	-1.33***	
	(.23)	
Observations	813	813

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Table G5. Binomial Logit Model PSA Scaled Score & NCA by Gender

	NCA	NCA (ME)
NCA scaled score	.03	
	(.09)	
Days b/w Release & Close	-.00*	
	(.00)	
Female	-.45	
	(.35)	
Sex: Male (ME)		.00
		(.01)
Sex: Female (ME)		.00
		(.01)
Constant	-1.87***	
	(.37)	
Observations	813	813

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Table G6. Binomial Logit Model NVCA Flag & NVCA Outcome by Gender

	NVCA	NVCA (ME)
NVCA flag	.00	
	(.)	
Days b/w Release & Close	-.01	
	(.01)	
Female	.62	
	(.85)	
Sex: Male (ME)		.00
		(.)
Sex: Female (ME)		.00
		(.)
Constant	-3.50***	
	(.61)	
Observations	793	793

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Appendix H: Classification, by Outcome & Defendant Characteristics

Table H1. FTA Classification, by Race

	N	High Risk	%	Accuracy (Correct Classification True Failures & Non-Failures)	Positive Predictive Value (Precision of Failure Prediction)	False Negative Rate (Miss Future Fail)	Negative Predictive Value (Precision of No Failure Prediction)	False Positive Rate (False Alarm)
Overall	**990**	**406**	**41%**	**59%**	**33%**	**38%**	**77%**	**50%**
White	265	141	53%	51%	34%	52%	70%	44%
Black or African American	394	147	37%	61%	31%	34%	78.9%	54%
Hispanic	276	102	37%	63%	36%	32%	78.2%	51%
"Other" Identity	55	16	29%	67%	38%	24%	79.5%	57%

Note: Cells with the highest rate across racial groups are filled with orange and cells with the lowest rate are filled with blue; Defendants classified as "high risk" for FTA had a FTA scaled score of at least 3.

Table H2. NCA Classification, by Race

	N	High Risk	%	Accuracy (Correct Classification True Failures & Non-Failures)	Positive Predictive Value (Precision of Failure Prediction)	False Negative Rate (Miss Future Fail)	Negative Predictive Value (Precision of No Failure Prediction)	False Positive Rate (False Alarm)
Overall	**990**	**464**	**47%**	**54%**	**10%**	**46%**	**92%**	**46%**
White	265	149	56%	43%	5%	57%	93%	53%
Black or African American	394	195	49%	53%	12%	48%	92%	39%
Hispanic	276	102	37%	63%	15%	35%	91%	50%
"Other" Identity	55	18	33%	67%	17%	31%	92%	50%

Note: Cells with the highest rate across racial groups are filled with orange and cells with the lowest rate are filled with blue; Defendants classified as "high risk" for NCA had an NCA scaled score of at least 3.

Table H3. NVCA Classification, by Race

	N	High Risk	%	Accuracy (Correct Classification True Failures & Non-Failures	Positive Predictive Value (Precision of Failure Prediction)	False Negative Rate (Miss Future Fail)	Negative Predictive Value (Precision of No Failure Prediction)	False Positive Rate (False Alarm)
Overall	**990**	**30**	**3%**	**96.3%**	**0.0%**	**3.1%**	**99.3%**	**100.0%**
White	265	10	4%	95.8%	0.0%	3.8%	99.6%	100.0%
Black or African American	394	16	4%	94.9%	0.0%	4.1%	98.9%	100.0%
Hispanic	276	4	1%	97.8%	0.0%	1.5%	99.3%	100.0%
"Other" Identity	55	0	0%	100.0%	[Omitted]	0.0%	100.0%	[Omitted]

Note: Cells with the highest rate across racial groups are filled with orange and cells with the lowest rate are filled with blue; Defendants classified as "high risk" for NVCA had an NVCA flag of "yes."

Appendix I: FTA & Violations Sensitivity Analysis

Table I1. Binomial Logit Model PSA Scaled Score & FTA + Violations

	FTA + Violations	*FTA + Violations (ME)*
FTA scaled score	.24***	.05***
	(.05)	(.01)
Days between release & close	-.00**	
	(.00)	
Constant	-1.07***	
	(.21)	
Observations	814	814

Note. PSA = Public Safety Assessment; FTA = failure to appear; ME = marginal effects.
Standard errors in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table I2. Binomial Logit Model PSA Scaled Score & FTA + Violations, by Race

	FTA + Violations	*FTA + Violations (ME)*
FTA scaled score	.23***	
	(.05)	
Days between release & close	-.00***	
	(.00)	
Race: Hispanic	.00	
	(.20)	
Race: Black	-.36	
	(.19)	
Race: Other	-.20	
	(.36)	
Race: White (ME)		.05***
		(.01)
Race: Hispanic (ME)		.05***
		(.01)
Race: Black (ME)		.05***
		(.01)
Race: Other (ME)		.05***
		(.01)
Constant	-.90***	
	(.26)	
Observations	814	814

Note. PSA = Public Safety Assessment; FTA = failure to appear; ME = marginal effects.
Standard errors in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table I3. Binomial Logit Model PSA Scaled Score & FTA + Violations, by Sex

	FTA + Violations	*FTA + Violations (ME)*
FTA scaled score	.24***	
	(.05)	
Days between release & close	-.00***	
	(.00)	
Sex: Female	.13	
	(.20)	
Sex: Male (ME)		.05***
		(.01)
Sex: Female (ME)		.05***
		(.01)
Constant	-1.09***	
	(.22)	
Observations	813	813

Note. PSA = Public Safety Assessment; FTA = failure to appear; ME = marginal effects.
Standard errors in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$

Appendix J: Supervised Population Sensitivity Analysis

Table J1. Binomial Logit Model PSA Scaled Score & FTA for Supervised Population

	FTA	*FTA (ME)*
FTA scaled score	.22***	.04***
	(.06)	(.01)
Days between release & close	-.00**	
	(.00)	
Constant	-1.36***	
	(.26)	
Observations	685	685

Note. PSA = Public Safety Assessment; FTA = failure to appear; ME = marginal effects.
Standard errors in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table J2. Binomial Logit Model PSA Scaled Score & NCA for Supervised Population

	NCA	*NCA (ME)*
NCA scaled score	.06	.01
	(.10)	(.01)
Days between release & close	-.00	
	(.00)	
Constant	-2.01***	
	(.40)	
Observations	685	685

Note. PSA = Public Safety Assessment; NCA = new criminal activity; ME = marginal effects.
Standard errors in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table J3. Binomial Logit Model NVCA Flag & NVCA Outcomes for Supervised Population

	NVCA	*NVCA (ME)*
NVCA flag	.00	.00
	(.)	(.)
Days between release & close	-.01	
	(.01)	
Constant	-3.79***	
	(.65)	
Observations	669	669

Note. PSA = Public Safety Assessment; NCVA = new criminal violent activity; ME = marginal effects.
Standard errors in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table J4. Binomial Logit Model PSA Scaled Score & FTA for Supervised Population, by Race

	FTA	***FTA (ME)***
FTA scaled score	.22***	
	(.06)	
Days between release & close	-.00**	
	(.00)	
Race: Hispanic	.06	
	(.23)	
Race: Black	-.23	
	(.22)	
Race: Other	.09	
	(.40)	
Race: White (ME)		.04***
		(.01)
Race: Hispanic (ME)		.04***
		(.01)
Race: Black (ME)		.04***
		(.01)
Race: Other (ME)		.04**
		(.01)
Constant	-1.27***	
	(.31)	
Observations	685	685

Note. PSA = Public Safety Assessment; FTA = failure to appear; ME = marginal effects.
Standard errors in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table J5. Binomial Logit Model PSA Scaled Score & NCA for Supervised Population, by Race

	NCA	*NCA (ME)*
NCA scaled score	.08 (.10)	
Days between release & close	-.00 (.00)	
Race: Hispanic	.81* (.36)	
Race: Black	.62 (.34)	
Race: Other	.76 (.56)	
Race: White (ME)		.01 (.01)
Race: Hispanic (ME)		.01 (.01)
Race: Black (ME)		.01 (.01)
Race: Other (ME)		.01 (.01)
Constant	-2.66*** (.50)	
Observations	685	685

Note. PSA = Public Safety Assessment; NCA = new criminal activity; ME = marginal effects. Standard errors in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table J6. Binomial Logit Model NVCA Flag & NCVA Outcome for Supervised Population, by Race

	NCVA	*NCVA (ME)*
NCVA Flag	.00 (.)	
Days between release & close	-.01 (.01)	
Race: Hispanic	-.10 (.92)	
Race: Black	.00 (.)	
Race: Other	.00 (.)	
Race: White (ME)		.00 (.)
Race: Hispanic (ME)		.00 (.)
Race: Black (ME)		.00 (.)
Race: Other (ME)		.00 (.)
Constant	-3.42*** (.75)	
Observations	458	458

Note. PSA = Public Safety Assessment; NCVA = new criminal violent activity; ME = marginal effects. Standard errors in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table J7. Binomial Logit Model PSA Scaled Score & FTA for Supervision Population, by Sex

	FTA	*FTA (ME)*
FTA scaled score	.22***	
	(.06)	
Days between release & close	-.00**	
	(.00)	
Sex: Female	.14	
	(.23)	
Sex: Male (ME)		.04***
		(.01)
Sex: Female (ME)		.04***
		(.01)
Constant	-1.37***	
	(.26)	
Observations	684	684

Note. PSA = Public Safety Assessment; FTA = failure to appear; ME = marginal effects.
Standard errors in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table J8. Binomial Logit Model PSA Scaled Score & NCA for Supervision Population, by Sex

	NCA	*NCA (ME)*
NCA scaled score	.05	
	(.10)	
Days between release & close	-.00	
	(.00)	
Sex: Female	-.46	
	(.37)	
Sex: Male (ME)		.00
		(.01)
Sex: Female (ME)		.00
		(.01)
Constant	-1.91***	
	(.40)	
Observations	684	684

Note. PSA = Public Safety Assessment; NCA = new criminal activity; ME = marginal effects.
Standard errors in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table J9. Binomial Logit Model NVCA Flag & NCVA Outcome for Supervision Population, by Sex

	NCVA	*NCVA (ME)*
NCVA flag	.00 (.)	
Days between release & close	-.01 (.01)	
Sex: Female	.19 (1.13)	
Sex: Male (ME)		.00 (.)
Sex: Female (ME)		.00 (.)
Constant	-3.83*** (.69)	
Observations	668	668

Note. PSA = Public Safety Assessment; NCVA = new criminal violent activity; ME = marginal effects. Standard errors in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$