FRESNO
PSA
VALIDATION
Initial Report

JSP
JUSTICE SYSTEM PARTNERS

Table of Contents

Site Description

Fresno County, California, is located in California's central valley region, where it is surrounded by national parkland and large-scale agricultural areas. The county seat is the city of Fresno, which is a major city with approximately 542,000 residents. As of 2020, Fresno is one of the largest majority-Latine areas in the United States, with 53.6% of the population categorized as Hispanic or Latino (any race) on the 2020 US Census. The median household income is $69,571, with 18.7% of all people living in poverty. Poverty is concentrated among children: 25.7% of people under 18 are in poverty compared to 16.6% of those aged 18-64.

Fresno County is served by both the Fresno County Sheriff, which provides court protection, jail administration, and coroner services, and by municipal police departments including the Fresno Police Department. The Sheriff operates the Fresno County Jail in downtown Fresno. Like many US localities, Fresno has seen recent reductions in violent crime, including decreases in homicide, gang-related shootings, burglaries and robberies[1]. Fresno Police Chief Paco Balderrama has acknowledged that domestic violence cases have increased and has recently hired more staff for the domestic violence unit[2]. Fresno has also made progress in lowering the number of drug overdose deaths, but rates remain high. 17% of all deaths investigated by the coroner in 2021 were due to drug overdose, predominantly methamphetamine (present in 61% of single drug overdose deaths) and fentanyl (present in 25% of single drug overdose deaths)[3].

Fresno Pretrial Services

The Fresno pretrial division is located within the Fresno County Probation Department. The purpose of the pretrial office is to promote due process to those accused of crime(s), to assist the Fresno Superior Court in making fair, prompt and effective release decisions at the Arraignment Hearing, maintain the integrity of the judicial process by ensuring individuals appear for court, protect victims and witnesses and provide community safety.

Pretrial Services completes a pretrial assessment and report utilizing the Public Safety Assessment (PSA) to assist judicial officers in identifying persons who are unlikely to accrue new arrests and have a high likelihood of appearing for all court appointments, and therefore can be released into the community pending resolution of their Court case(s).

---

[1] https://abc30.com/crime-rates-nationwide-trend-violence-less-homicides/14256847/

[2] https://gvwire.com/2023/12/31/fresno-police-say-major-crime-was-down-in-2023/

[3] https://www.fresnosheriff.org/images/pdfs/2021_Coroner_Unit_Statistics.pdf

The Court then decides whether the person is suitable for Pretrial Release and if any conditions should be imposed.

Fresno County began implementation of the PSA in the Fall of 2021 which included stakeholder training.   Pretrial staff were trained in 2021 and they began administering the assessment on November 1, 2021.

PSA Instrument Description

The Public Safety Assessment (PSA) is an actuarial assessment developed with funding from the John and Laura Arnold Foundation (now Arnold Ventures) to help jurisdictions assess the risk of failure to appear in court for pretrial hearings, any new criminal arrests, and new violent criminal arrests for individuals on pretrial release. The PSA was designed to provide decisionmakers with research-supported information that can be used when making release decisions and setting release conditions.

The PSA is administered through a series of administrative data reviews (based on previous criminal history, failure to appear, and factors associated with the current offense) in which three scales are produced: FTA, NCA, and NVCA. Each scale is comprised of individual items scored and summed and then plotted on the scale.

The **Failure to Appear (FTA) scale** is designed to predict the likelihood that people will return to court (versus "fail to appear") for pretrial hearings. The **FTA Scale** is comprised of 5 individual items that total seven points. Three of the items (Pending charge at time of the offense, prior conviction, and prior failure to appear older than 2 years) are scored 0 or 1 while the fourth item, Prior failure to appear pretrial within the past 2 years is scored a 0 for no priors, 2 points for 1 prior FTA, and 4 points for 2 or more prior FTAs within the past two years. Once the raw score is tabulated, it is then applied to the FTA Scale and a subsequent "score" is produced.

The **New Criminal Arrest (NCA) scale** provides decisionmakers with information regarding the likelihood that an individual will be arrested for a new crime while on pretrial release. The **NCA scale** is also plotted on a six-point scale that is comprised of up to 13 points for the individual raw score. The first item, age at current arrest is scored as a 0 for individuals who were 23 or older while those that are 22 or younger at the time of their arrest are scored a 2. Prior misdemeanor and prior felony convictions are each scored as 0 and 1, with one being any prior conviction. If a person has a pending charge at the time of the new arrest, they are scored a 3 while those with a prior sentence to incarceration is scored a 2. The sixth item, prior violent conviction is scored as a 0 for no priors, a 1 for 1 or 2 priors, and a 2 for 3 or more prior violent convictions. And the seventh and final item for the NCA scale

is based on the number of prior failures to appear within the past two years. If the person has no prior FTA in the past year, a score of 0 is given, 1 prior FTA results in a score of 1, while 2 or more prior FTA is scored as a 2. Once tabulated and summed, the final score is plotted on the NCA scale and a subsequent "scale score" is produced.

In addition to the FTA and NCA scales, the PSA also provides a score associated with **New Violent Criminal Activity (NVCA)**. The NVCA raw score ranges from 0 to 7 across 5 items. Current violent offense + 20 years old or younger, pending charge at the time of the offense, and prior conviction (for any reason) are all scored 0 (no) and 1 (yes). If the person's current offense is violent, they receive a score of 2 for this item. And for the last item, prior violent conviction a person with no prior violent convictions is scored a 0, a person with 1 or 2 prior violent convictions a 1, and 3 or more violent convictions a 2. Unlike the other scales, the **NVCA is designed as a violent flag**—with people who score a 3 or less identified as not having the violence flag and those scoring 4 or more are identified with a violence flag.

In the table below, gray boxes indicate that the risk factor item is included in that scale. For example, the FTA scale includes items 1, 2, 2a, 4, 5, 6, and 9. The item scores for the FTA and NCA scales are weighted to produce an individual scale score of 1-6. The NVCA items are weighted to form a yes/no outcome.

	Scales		
	FTA	**NCA**	**NVCA**
Risk Factor			
1. Age at current offense		█	
2. Current violent offense			█
2a. Current violent offense and ≤20 years			█
3. Pending charge at time of offense	█	█	█
4. Prior misdemeanor conviction		█	
5. Prior felony conviction		█	
5a. Prior conviction	█		█
6. Prior violent conviction (incl. #)		█	█
7. Prior FTA pretrial in past 2 years (incl. #)	█	█	
8. Prior FTA pretrial older than 2 years	█		
9. Prior sentence to incarceration		█	
*Boxes shaded blue indicate that the item is included in the scale; boxes shaded gray indicate that the item is **not** included in the scale.*			

Validation Method

Data Analysis

Sample

There were 14,612 assessments completed from November 2021 to August 2023. Of those, 11,488 were completed between July 2022 and August 2023 for anyone in the Fresno County Jail booked for an open charge (including warrants) except for those remanded by the Court, those with a Federal Hold, and those set for sentencing. 3,124 assessments were completed from November 2021 through June 2022. During this time, staffing issues necessitated more selective criteria for assessment. PSAs were completed for anyone in the Fresno County Jail booked for an open charge (including felony warrants) but did not review defendants having only misdemeanor warrants or defendants on any form of formal supervision.

To be included in this validation, cases must be resolved and the individual must have been released from jail at any point prior to disposition. Of all 14,612 PSAs completed, 3,978 cases (27.2%) met the criteria for the validation sample.

Sensitivity and Diagnostic Analyses

In the process of cleaning and preparing the data for analysis, we observed 570 (14.3%) individuals with the same identification number but variations in charges and filing dates. This is because the data collection tracked individual PSAs, not bookings or individual cases. Several PSAs might originate from one booking if, for example, cases are assigned to different courtrooms. PSA assessments are completed on warrants, so multiple PSA scores might originate from the same case over time if the individual has a warrant issued due to failure to appear.

To avoid double-counting individuals in our analysis, we identified these multiple cases and retained only the cases with the most recent filing date. Additionally, we found one case where an error in data entry resulted in the values being shifted one column to the left. This was fixed manually to allow for inclusion of this case in the analysis. The result was a total of 3,408 cases.

Sensitivity analyses suggest that this did not meaningfully change the outcome of our statistical tests – that is, we received similar validation results whether we included all cases or removed the duplicate cases, suggesting that the removal of the duplicates did not change our conclusions about the PSA tool's performance.

Outcome Measures

After cleaning and organizing the data, we conducted a series of analyses examining the relationships between each of the three PSA scales: Failure to Appear (FTA), New Criminal Arrest (NCA), and New Violent Criminal Arrest (NVCA).

The definition of "violent" crime varies between jurisdictions. In this case, violent crimes are defined in [California Penal Code PC 667.5(c)](#).

Table 1, Fresno Validation Outcome Measures, Definitions, and Values				
Outcome	*Abbreviation*	*Definition*	*"0" Value*	*"1" Value*
Failure to Appear	*FTA*	Any failure to appear	Did not fail to appear	One or more failures to appear
New Criminal Arrest	*NCA*	Any new criminal arrest under pretrial release	Did not have a new arrest while released pretrial	Had one or more new arrests while released pretrial
New Criminal Violent Arrest	*NVCA*	One of more new arrests include a violent arrest	No new arrests or new arrests do not include a violent arrest	One or more new arrests were for a violent arrest while released pretrial

For the FTA outcome, the outcome of interest is *any failure to appear*, where "0" indicates the individual did not fail to appear and "1" indicates the individual had one or more failures to appear. Similarly, the outcome of interest for the NCA scale is defined as *any new criminal arrests* while under pretrial release, and this was also coded as "0" indicating the individual had no new arrests while on pretrial release and "1" indicates the individual has one or more new arrests while under pretrial release. If the individual had at least one new arrest, a separate variable indicates how many of those new arrests were considered violent crimes (range: 0-4). The New Violent Criminal Arrest (NVCA) outcome is defined as *one or more new arrests include a violent arrest*. To measure this, if any of the new arrests were not for a violent offense, this was coded as "0." If the individual had no new arrests while under pretrial release, this was also coded as "0". If the individual had a new arrest and *any* of their new arrests included a violent charge, then it was recoded as "1."

ROC Analyses

Receiver Operating Characteristics/Area Under the Curve (ROC/AUC) is used to determine the probability that a randomly selected individual from the data who failed (new arrest, new arrest for violent offense, or FTA) had a higher score than a randomly selected

7

individual that did not fail. If the instrument (in this case the PSA) fails to identify any differences between individuals who failed and individuals who did not, then the AUC score is .500.

While there is no exact AUC score that suggests that an assessment is valid, the following is used as a general guide or best practice for interpreting the results of ROC/AUC analyses:

Table 2, ROC/AUC Scores and Interpretations	
AUC Score	Interpretation
.54 or less	No evidence of validity
.55 to .63	Weak evidence
.64 to .70	Moderate evidence
.71 and above	Strong evidence

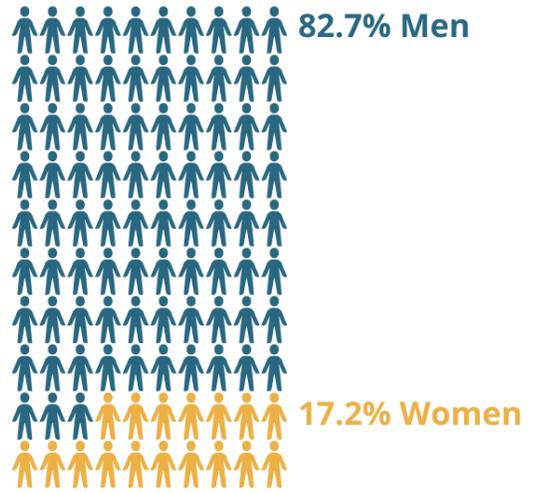## Sample Descriptives and Case Characteristics

The following pages contain descriptions of the sample for this analysis, including demographic measures and characteristics of the cases (original charge, disposition, etc). This information can be found in tables in Appendix A and B.
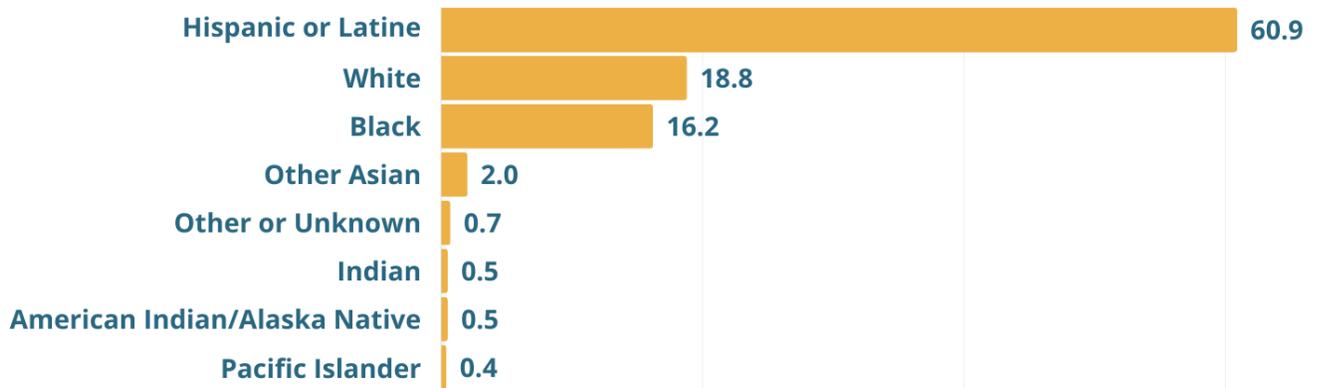
Sample Characteristics

Age at First Offense

- 20 or younger, 8.1%
- 21 or 22, 6.5%
- 23 or older, 85.4%

Gender[4, 5]

82.7% Men

17.2% Women

Race and Ethnicity[6]

Hispanic or Latine	60.9
White	18.8
Black	16.2
Other Asian	2.0
Other or Unknown	0.7
Indian	0.5
American Indian/Alaska Native	0.5
Pacific Islander	0.4
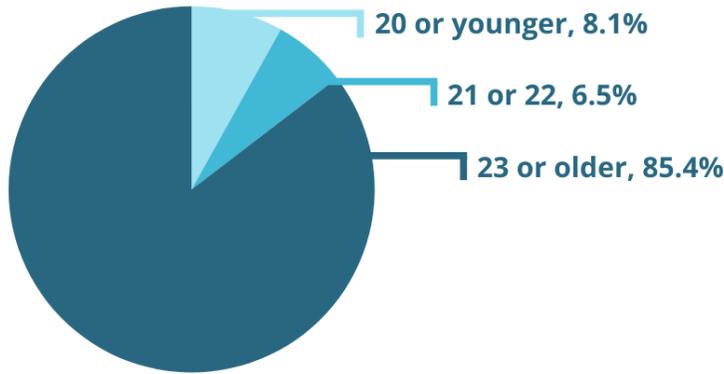
Age at Booking

34.6 years

Mean Age at Booking
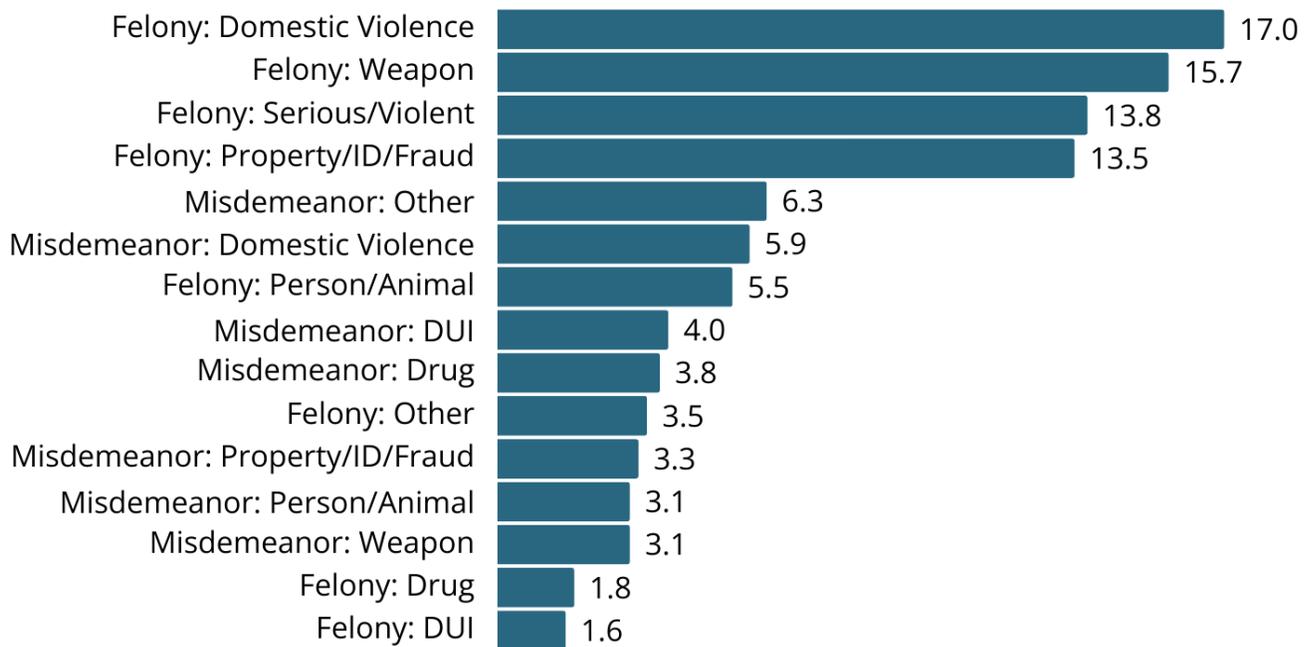
18 → 92

Age Range of Sample

---

[4] We use the term "sex" to present the categorization as captured by the agency (which uses the variable name "Gender" but categories of M, F, and T; it is not clear whether this variable is based on self-reported, perceived, or assigned identity). We recognize the demographic representation of our sample may be different when considering gender identity and could include representation from individuals who identify as non-binary or gender non-conforming. We believe it is important for administrative data to consider the full spectrum of identity to understand the individual experience and disparate outcomes more adequately. This includes disaggregating the "transgender" category into transgender women, transgender men, and transgender nonbinary or gender non-conforming people.

[5] One case was identified as a transgender person. The case could not be included in the validation analysis because a single case is not sufficient for drawing conclusions about the PSA's performance for transgender individuals.

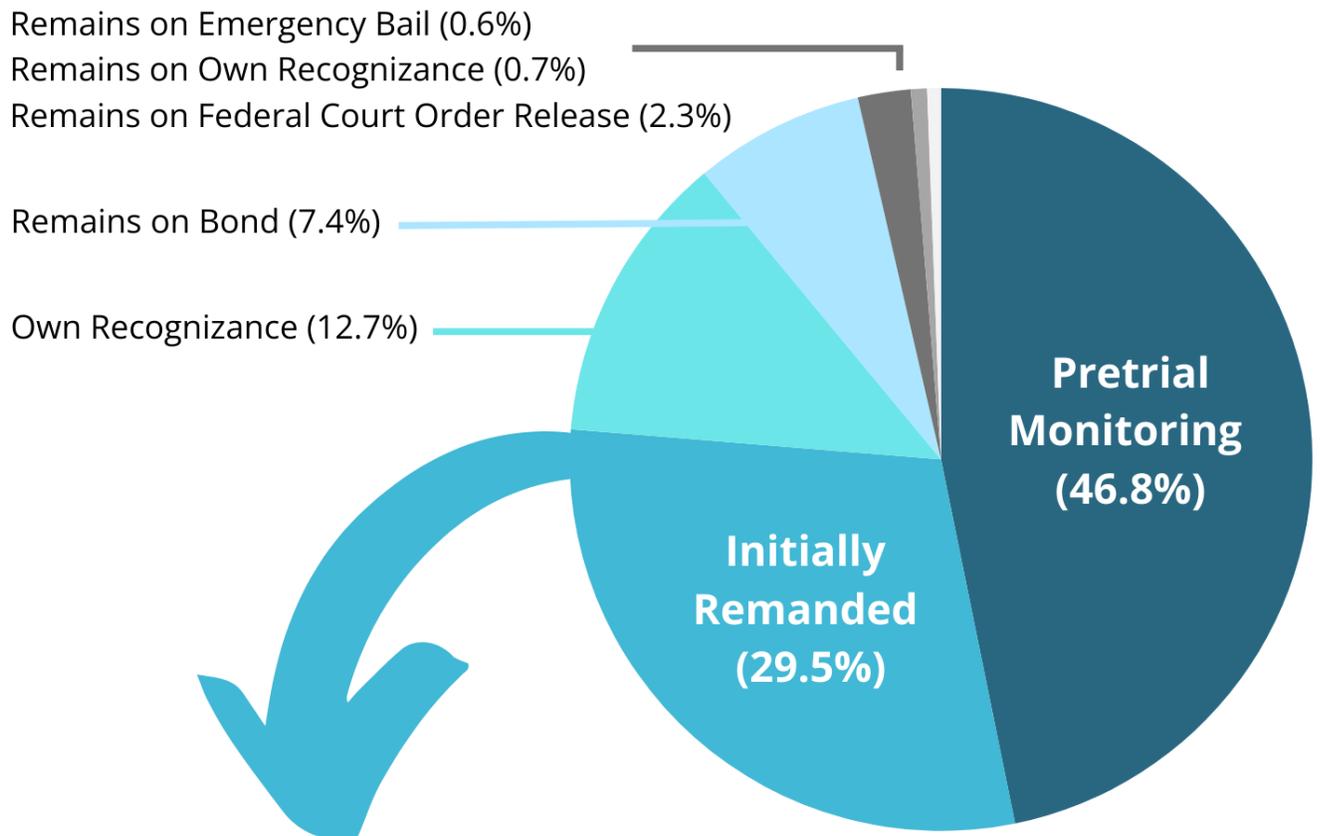[6] We report race as the mutually exclusive categories captured by the agency. We cannot determine individuals in the sample who identify as bi- or multiracial, or who identify in other ways than what is captured by the agency. We recognize self-reported racial identity is critical for accurately reporting the true demographic profile of the sample, the individual's experience, and any disparities.

Case Characteristics

Original Charge Types

Charge Type	Value
Felony: Domestic Violence	17.0
Felony: Weapon	15.7
Felony: Serious/Violent	13.8
Felony: Property/ID/Fraud	13.5
Misdemeanor: Other	6.3
Misdemeanor: Domestic Violence	5.9
Felony: Person/Animal	5.5
Misdemeanor: DUI	4.0
Misdemeanor: Drug	3.8
Felony: Other	3.5
Misdemeanor: Property/ID/Fraud	3.3
Misdemeanor: Person/Animal	3.1
Misdemeanor: Weapon	3.1
Felony: Drug	1.8
Felony: DUI	1.6

Pretrial Release Type

Remains on Emergency Bail (0.6%)
Remains on Own Recognizance (0.7%)
Remains on Federal Court Order Release (2.3%)

Remains on Bond (7.4%)

Own Recognizance (12.7%)

Pretrial Monitoring (46.8%)

Initially Remanded (29.5%)

Of those initially remanded...

Eventual Release Type

Release Type	Value
Bond	~45
Federal Court Order Release	~19
Own Recognizance	~17
Pretrial Monitoring	~15
Release to Program	~3
Emergency Bail After	~1

(Horizontal axis: 0, 10, 20, 30, 40, 50)
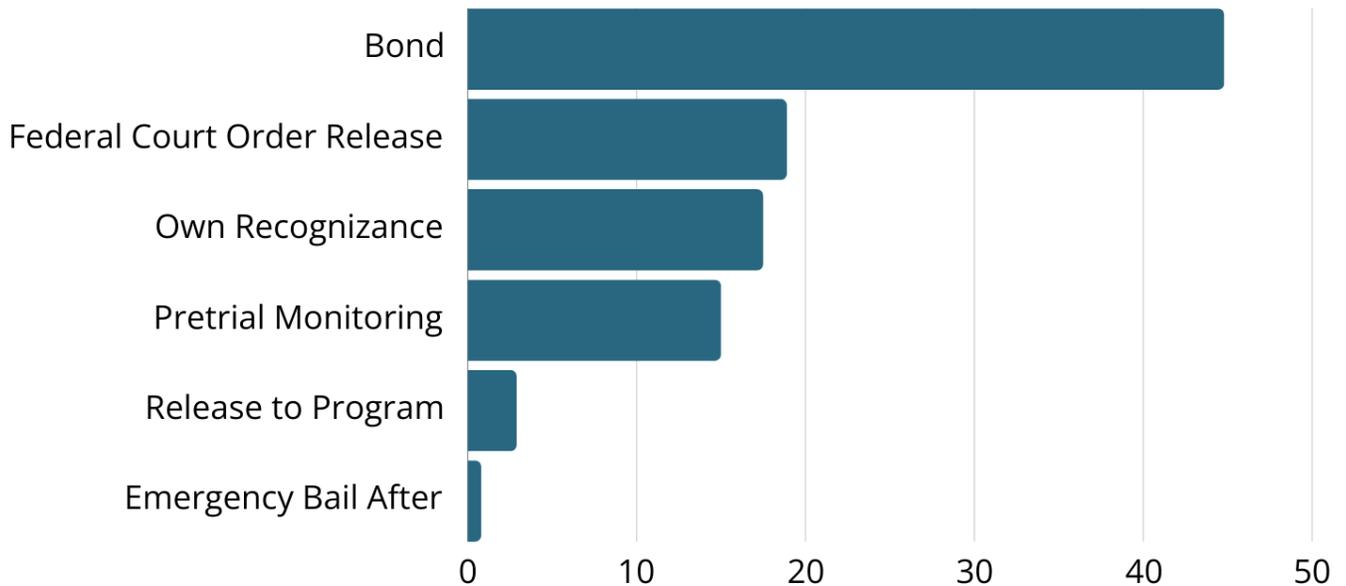
The average time between booking and release for those remanded was **34 days**, resulting in a total of **34,254 bed-days** across the observation period.

Case Disposition (Full Sample)

- Sentenced to Probation (50.3%)
- Dismissed (36.5%)
- Civil Commitment (0.9%)
- Sentenced to Local Prison (2.3%)
- Sentenced to Jail (4.3%)
- Sentenced to State Prison (5.6%)
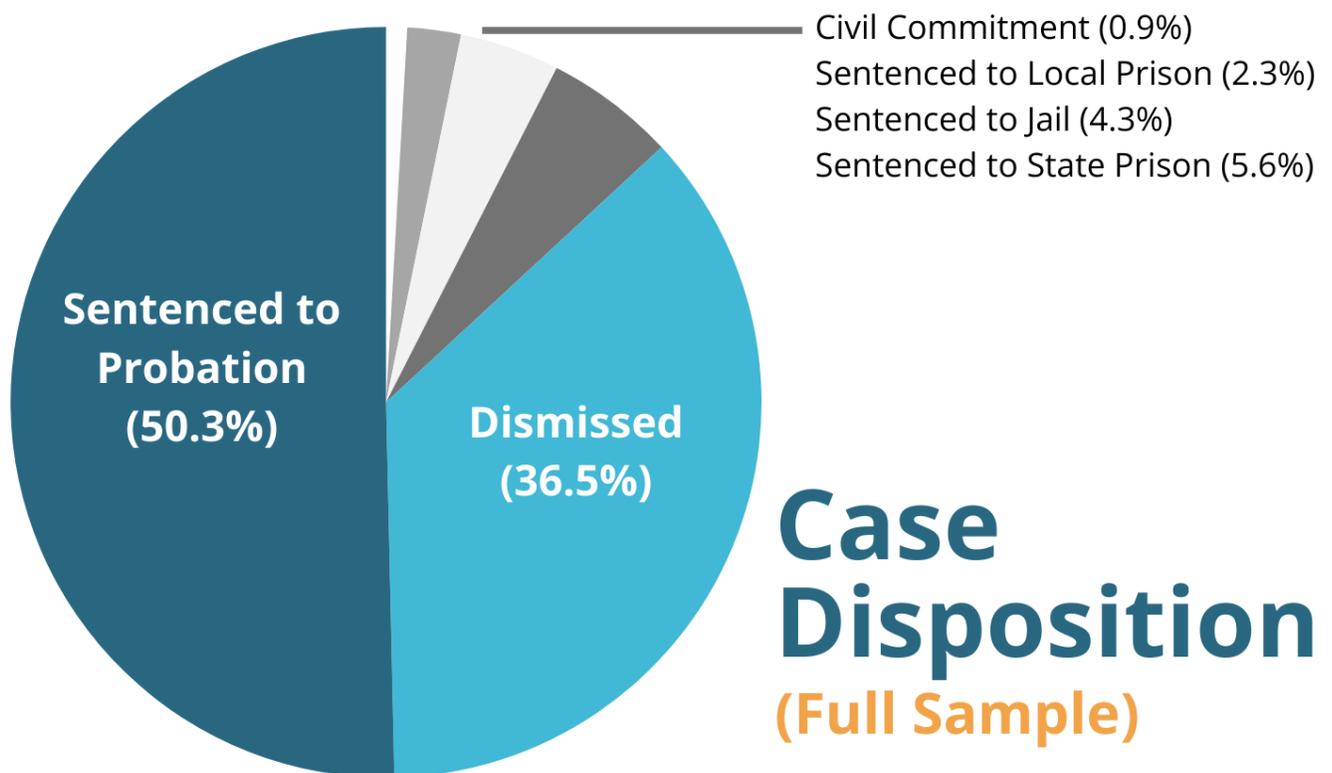
PSA Assessment Descriptives

As previously described, the PSA is comprised of 3 measures: Failure to Appear (FTA), New Criminal Arrest (NCA), and New Violent Criminal Arrest (NVCA). The tables and charts below show the distribution of scale scores across the full validation sample of 3,408 cases.

Scores for the items included in the FTA scale are summed and weighted into an FTA scale score, which can range from 1 (lowest likelihood of FTA) to 6 (highest likelihood of FTA).

FTA Scale Scores, Proportion of Total

19.5%	16.6%	20.5%	15.8%	15.9%	11.7%
1	2	3	4	5	6
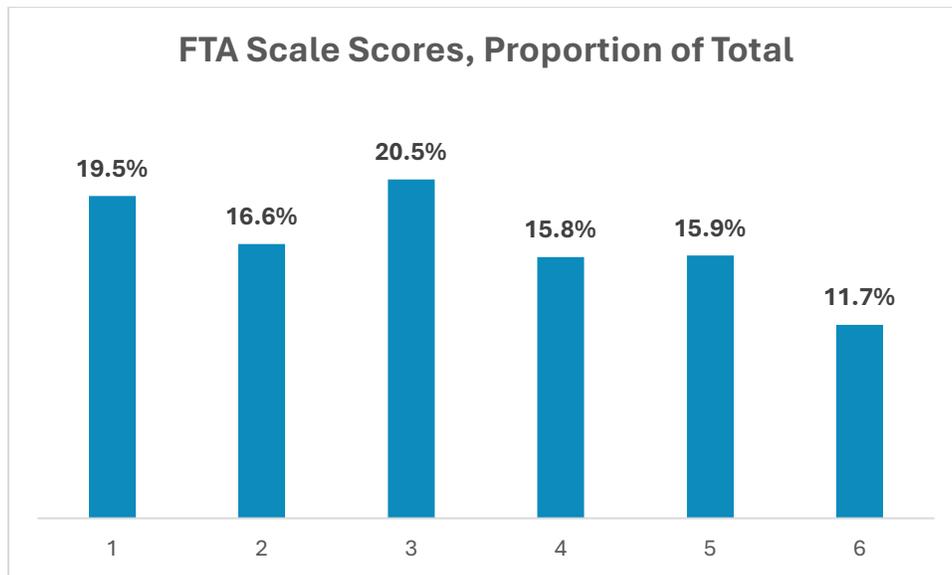
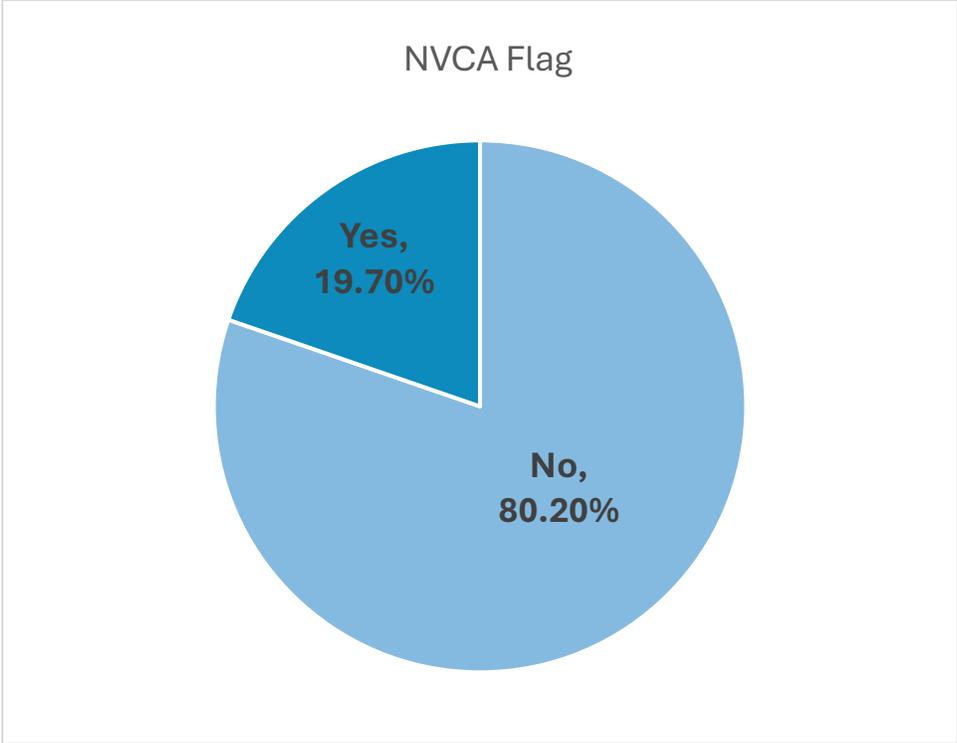| Table 3, FTA Scale Score for Sample		
	f	%
1	665	19.5%
2	567	16.6%
3	698	20.5%
4	538	15.8%
5	542	15.9%
6	398	11.7%
Total	3408	100%

Scores for the items included in the NCA scale are summed and weighted into an NCA scale score, which can range from 1 (lowest likelihood of NCA) to 6 (highest likelihood of NCA).

NCA Scale Scores, Proportion of Total

Score	Percentage
1	12.0%
2	16.3%
3	21.4%
4	24.9%
5	12.9%
6	12.5%
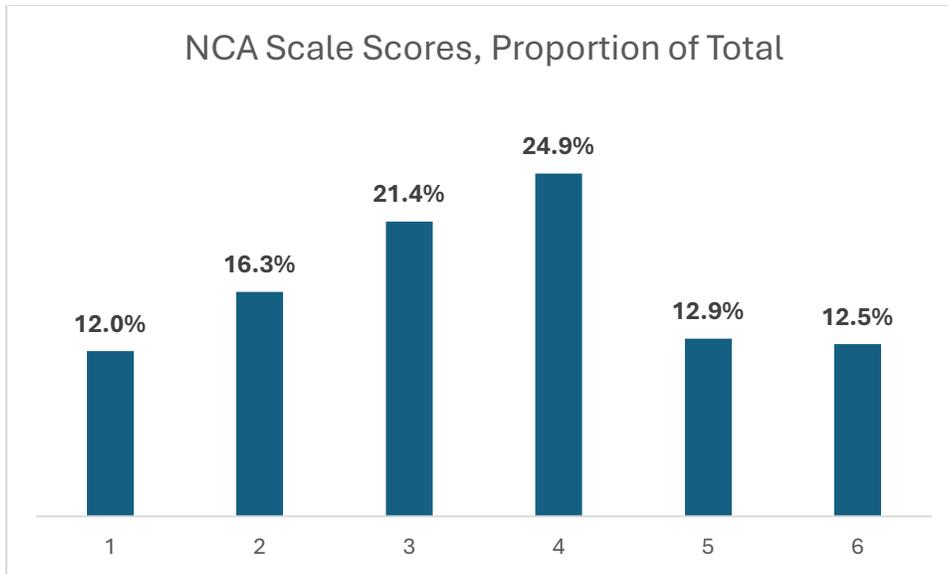
Table 4, NCA Scale Score for Sample		
	f	%
1	408	12%
2	555	16.3%
3	731	21.4%
4	848	24.9%
5	439	12.9%
6	426	12.5%
Total	3408	100%

The New Violent Criminal Arrest (NVCA) scale differs from the FTA and NCA scales in that it acts as a "flag," where raw scores of 0-3 receive a "No" (coded here as 0) and raw scores of 4 and above receive a "Yes" (coded as 1). This results in only two possible values instead of six like the FTA and NCA scales.

NVCA Flag

Table 5, NVCA Flag (Yes, No) for Sample		
	f	*%*
No	2734	80.2%
Yes	673	19.7%
Total	3408	100%

Bivariate Analyses of Demographics, Scale Scores and Outcome Measures

Failure to Appear Scale

Table 6, Bivariate Comparisons of FTA Scale Scores						
	FTA Scale Score					
	1	**2**	**3**	**4**	**5**	**6**
Sex						
Men	18.5	17.8	21	15.8	15.2	11.7
Women	24.4	11.2	17.9	15.8	19.3	11.4
Race						
White	13.4	15.6	23.1	15.3	18.4	14.3
Hispanic	21.2	15.2	19.6	16.5	15.6	11.9
Black	18.9	21.4	21.8	14.5	14.3	9.1
Pretrial Release Type						
Granted PT	51.0	49.0	52.6	42.8	45.2	34.4
OR	24.2	15.7	11.3	8.9	6.1	6.0
Remains on Bond	7.1	7.9	8.3	9.9	5.0	5.3
Remains on Emergency Bail	0.5	0.4	0.4	0.2	0.7	1.5
Remains on FCR	0.6	0.7	1.7	3.5	3.9	4.5
Remains on OR	0.9	09	0.7	0.6	0.9	0
Remand	15.8	25.4	24.9	34.2	38.2	48.2

The above table shows that, as intended, increasing FTA scores are associated with lower proportions of people granted PT or OR. As FTA scores increase, the proportion of people remanded to custody increases. We see a notable difference between white individuals and Black or Hispanic individuals in the likelihood of receiving an FTA score of 1.

Table 7, Bivariate Comparisons of NCA Scale Scores						
	NCA Scale Score					
	1	**2**	**3**	**4**	**5**	**6**
Sex						
Men	10.4	16.3	21.1	25.4	13.5	13.2
Women	19.3	16.2	23.0	22.5	9.9	9.2
Race						
White	10	11.1	24.6	23.4	15	16

Hispanic	12.2	18.5	20.1	24.9	12.5	11.9
Black	11.3	14.5	22.0	27.9	11.8	12.5
Court Outcome						
Granted PT	46.8	51.4	53.6	46.2	41.5	35.9
OR	27.0	21.6	11.9	7.8	5.9	5.9
Remains on Bond	9.3	6.5	8.1	7.0	8.4	5.2
Remains on Emergency Bail	0.5	0.2	0.7	0.4	0.5	1.4
Remains on FCR	0.5	0.9	1.8	2.6	3.4	4.9
Remains on OR	1.2	0.9	0.7	0.6	0.5	0.5
Remand	14.7	18.6	23.3	35.5	39.9	46.2

As expected, increasing NCA scale scores are associated with decreasing likelihood of being granted PT or OR, and increasing likelihood of being remanded. We see that men typically receive higher NCA scale scores than women, and that white individuals are more likely to receive higher NCA scale scores.

| Table 8, Bivariate Comparisons of NVCA Flag		
	NVCA Flag	
	No	Yes
Sex		
Men	78.9	21.1
Women	86.5	13.5
Race		
White	82.2	17.8
Hispanic	81.5	18.5
Black	71.1	28.9
Court Outcome		
Granted PT	81.4	18.5
OR	91.7	8.3
Remains on Bond	71.3	28.7
Remains on Emergency Bail	89.5	10.5
Remains on FCR	88.5	11.5
Remains on OR	79.2	20.8
Remand	80.2	19.7

The bivariate comparisons for the NVCA flag show that men are more likely to receive a "Yes" NVCA flag than women. There is also a large difference among races. The percentage of Black individuals receiving the "Yes" flag is 10% higher than the percentages for white

and Hispanic individuals. This may mean that Black individuals are less likely to be released on their own recognizance (OR).

PSA Scale Scores and Case Outcomes Analyses

Failure to Appear (FTA) Scores and Outcomes

To analyze whether the FTA scale score predicts any FTA outcome, we conducted an ANOVA test to compare the mean proportions of any FTA for each FTA scale score group.

FTA Scale Validation:
Score Groups x % with any FTA
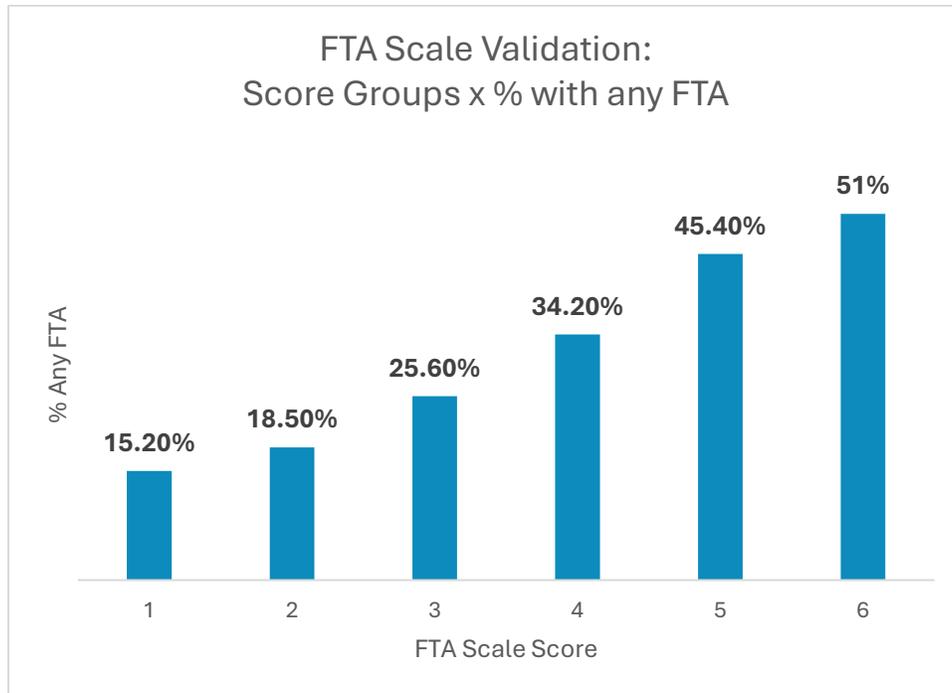
The ANOVA results show significant differences between groups (F = 56.5, p < .001). The effect size (η^2 = .077) indicates a moderately strong relationship between the FTA scale score and FTA outcome. The ROC AUC analysis produced a test result of .67, which is comparable to a moderate effect size (Rice & Harris, 2005).

We used a Sidak post hoc analysis to identify between-group differences. Individuals with FTA scores of 1 and 2 do not have significant differences in FTA outcomes but do differ from individuals with higher scores (4-6). At the higher end of the scale, individuals who score 5 or 6 are not significantly different from each other but are significantly different from individuals who score lower (1-4). The graphic below provides a visual representation of where we see differences between the score groups, with significant differences shaded in dark blue and no difference shaded in light blue.

Table 9, Significant Differences in FTA Outcomes Between FTA Scale Scores						
	1	2	3	4	5	6
1						
2						
3	*					
4	*	*	*			
5	*	*	*	*		
6	*	*	*	*		

Cells marked with a * indicate that these scale scores have significantly different FTA outcomes. For example, the likelihood of FTA is significantly different between those who score a 1 on the FTA scale and those who score a 3, but there is no significant difference in likelihood of FTA between individuals who score a 1 or a 2.

New Criminal Arrest (NCA) Scores and Outcomes

The following analysis compares the mean proportion of people at each NCA scale level with any new criminal arrest (NCA) in the observation period.

NCA Scale Validation:
Score Groups x % with any NCA
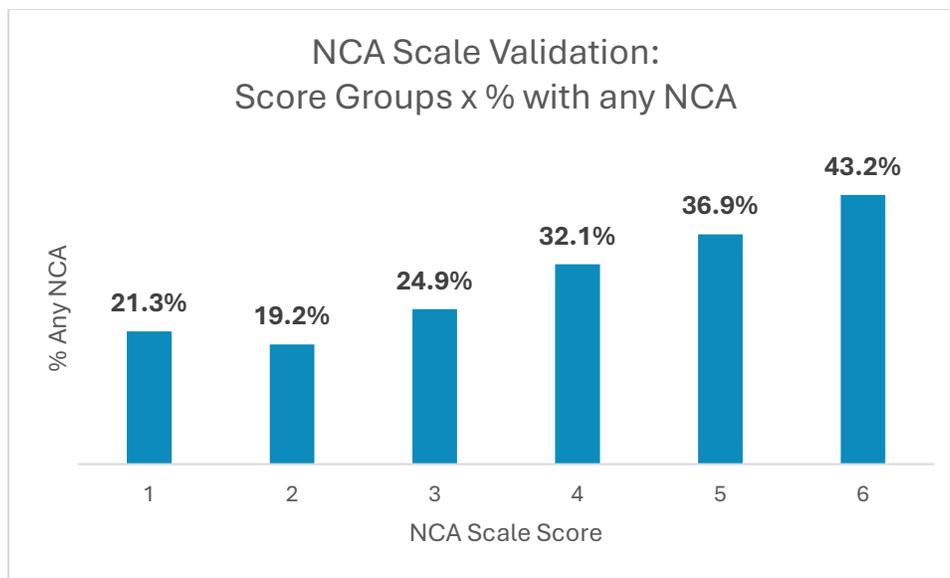
The ANOVA results show significant differences between groups (F = 20.954, p < .001). The effect size (η^2 = .03) indicates a small-to-moderate relationship between the NCA scale score and NCA outcome. The ROC/AUC analysis produced a test result of .61, which provides some weak evidence of validity.

The graphic below provides a visual representation of the Sidak post-hoc analysis, where we once again see significant differences between lower and higher scores, but failure to distinguish between neighboring scores (e.g., no difference between scores of 1 and 2 or between 5 and 6).

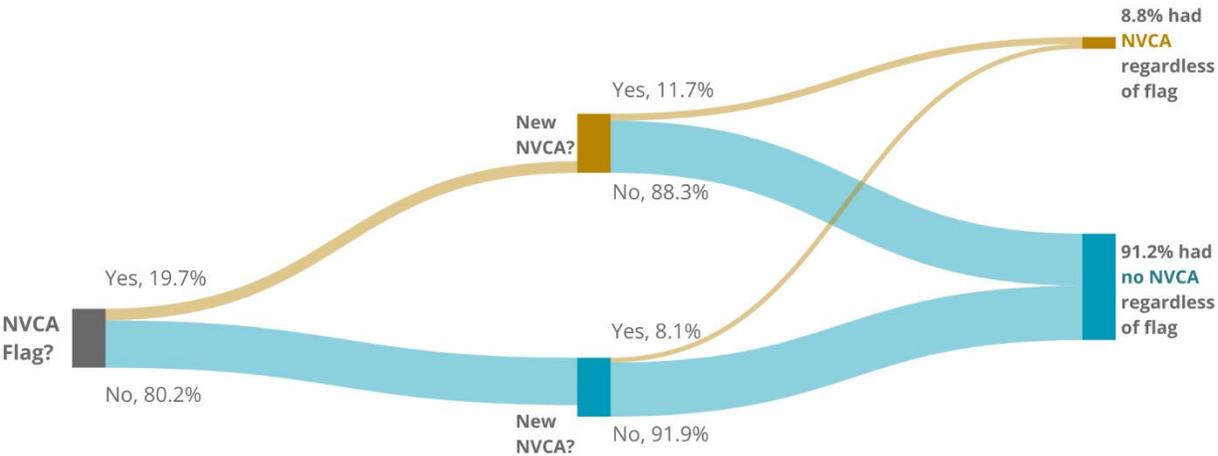| Table 10, Significant Differences in NCA Outcomes Between NCA Scale Scores						
	1	2	3	4	5	6
1						
2						
3						
4	*	*	*			
5	*	*	*			
6	*	*	*	*		

Cells marked with a * indicate that these scale scores have significantly different FTA outcomes. For example, individuals who score 1 on the NCA scale do not have significantly different likelihood of NCA than people who score a 2 or a 3, but are significantly different from people who score 4-6 (in this case, people who score a 1 have a significantly *lower* likelihood of NCA than people scoring 4-6).

New Violent Criminal Arrest (NVCA) Flag and Outcomes

The New Violent Criminal Arrest (NVCA) scale acts as a "flag," where raw scores of 0-3 receive a "No" (coded here as 0) and raw scores of 4 and above receive a "Yes" (coded as 1), and therefore an ANOVA is no longer appropriate. To validate the NVCA flag, we used a chi-square (χ^2) test of independence.

Presence of NVCA Flag and Any NVCA Outcome

- NVCA Flag?
  - Yes, 19.7%
  - No, 80.2%
- New NVCA?
  - Yes, 11.7%
  - No, 88.3%
- New NVCA?
  - Yes, 8.1%
  - No, 91.9%
- 8.8% had NVCA regardless of flag
- 91.2% had no NVCA regardless of flag

This graph shows that most people in the sample did not have new violent criminal arrest while under pretrial release, regardless of their NVCA score. 91.9% of those not flagged had no new violent arrests, and 88.3% of those flagged had no new violent arrests. Of those who were *not* flagged, 8.1% had a new violent criminal arrest while released pretrial. Of those who *were* flagged, 11.7% had a new violent criminal arrest. This presents a significant difference ($\chi^2 = 8.99$, p = .003), but with a very small effect size (Cramer's V = .05).
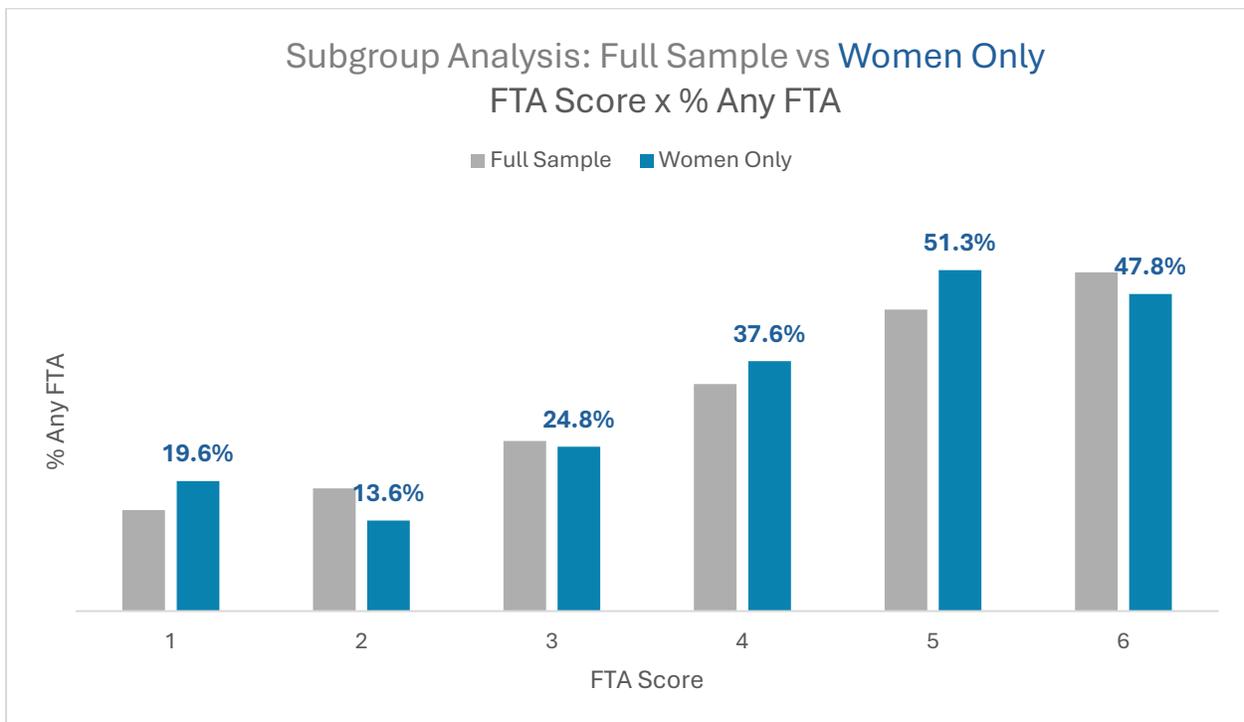
Subgroup Analyses

Sex

FTA Scores and Outcomes by Sex
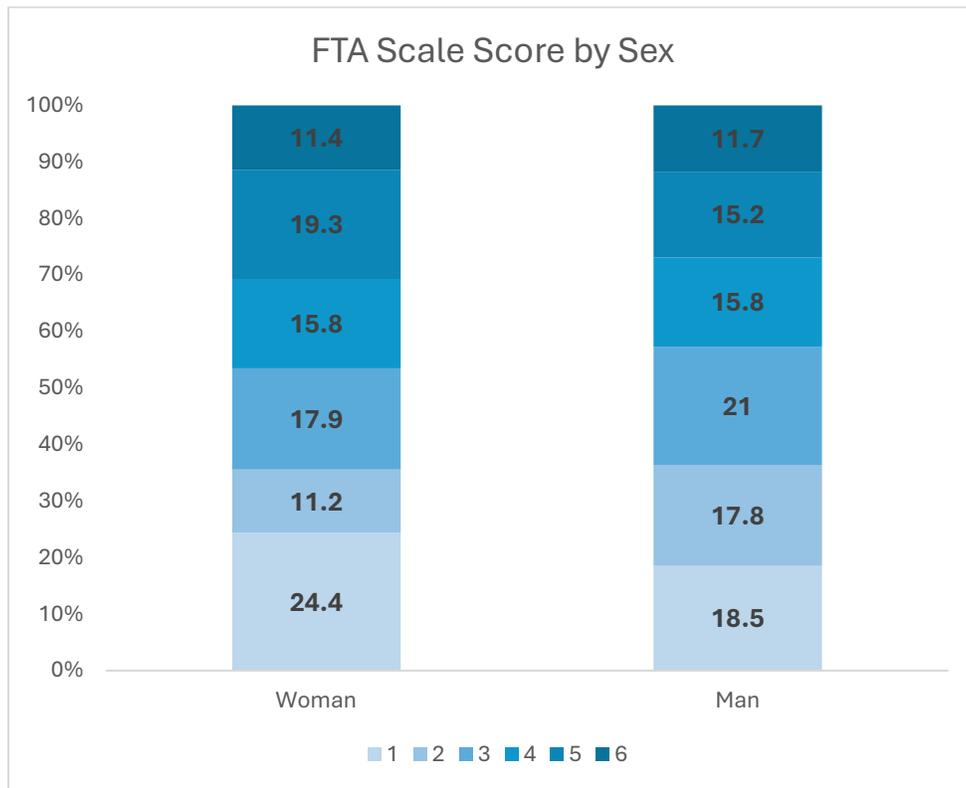
Proportionally, more women than men score a 1 or a 5 on the FTA scale ($\chi^2 = 32.54$, p < .001). Men are more likely than women to score a 2, 3, or 6. Some gender-neutral assessments are known to overclassify women, so the gender discrepancy at score 5 is worth further investigation.

FTA Outcome by Sex

Woman: No FTA 68.0%, Any FTA 32.0%
Man: No FTA 70.6%, Any FTA 29.4%
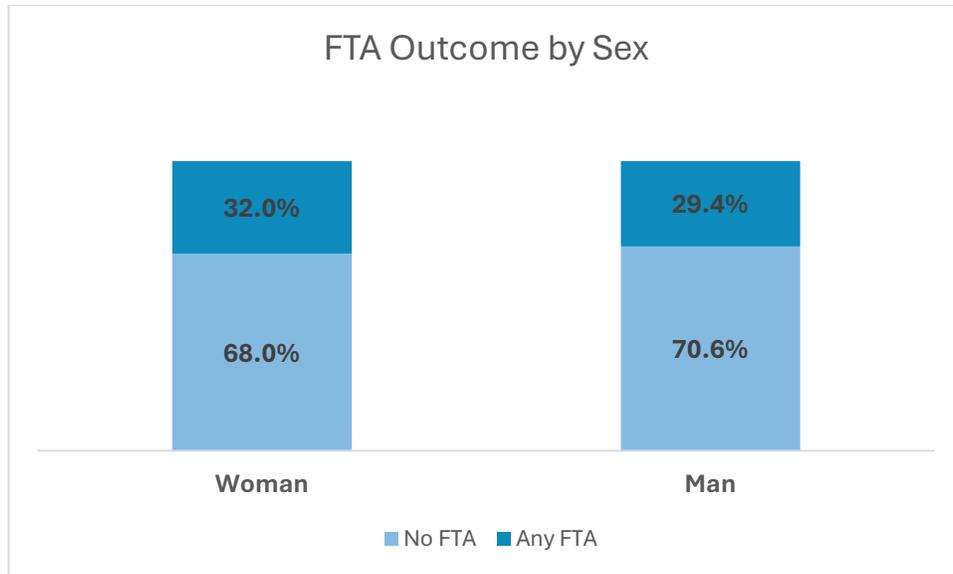
Women were slightly more likely to have any Failure to Appear incidents, but this is not a statistically significant difference (χ^2 = 1.99, p = .370). This suggests that the small differences between women's and men's FTA scale scores did not translate into meaningful differences in their actual likelihood of failing to appear for a pretrial hearing.

Subgroup Analysis: Full Sample vs Women Only
FTA Score x % Any FTA

Full Sample / Women Only

% Any FTA vs FTA Score

1: 19.6%
2: 13.6%
3: 24.8%
4: 37.6%
5: 51.3%
6: 47.8%

The subgroup analysis for women in the sample (n = 587) demonstrates the differences between the FTA scale validation results for the full sample, in gray, and for women only

(blue, labeled). In the above chart we see that women who scored a 1, 3, or 5 on the FTA scale were slightly more likely to fail to appear than the sample overall, whereas women who scored a 2, 3, or 6 were slightly less likely to have any FTA. The effect size (η^2 = .087) for women alone is slightly larger than that for the whole sample (η^2 = .077). The ROC/AUC analysis for the women-only subgroup resulted in an AUC of .67, which is equal to the ROC for the whole sample and indicates moderate evidence of validity.

NCA Scores and Outcomes by Sex
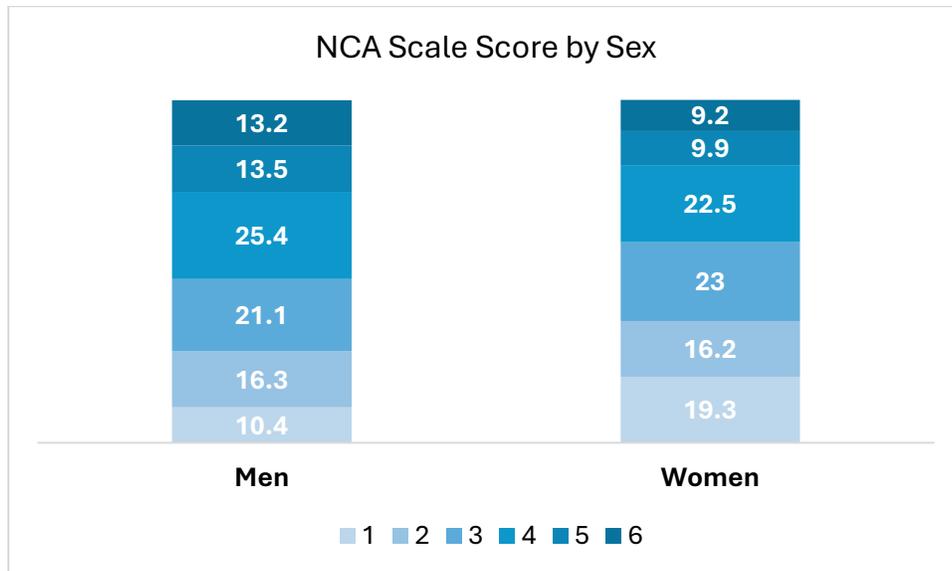
As shown above, there are significant (χ^2 = 52.6, p < .001) differences between sexes in NCA scale scores. Women were more likely to score on the lower end of the NCA scale (3 or below), and men were more likely to score on the higher end (4 or above).
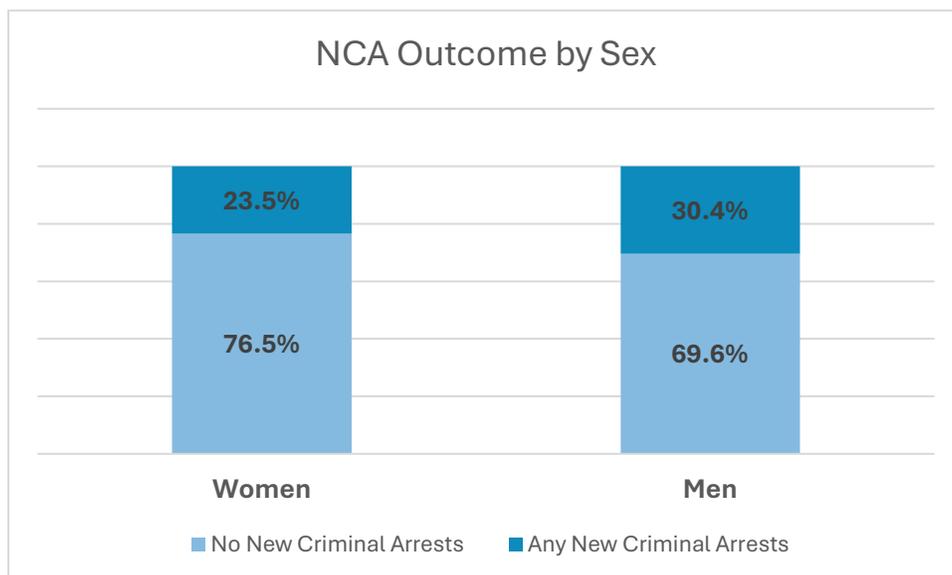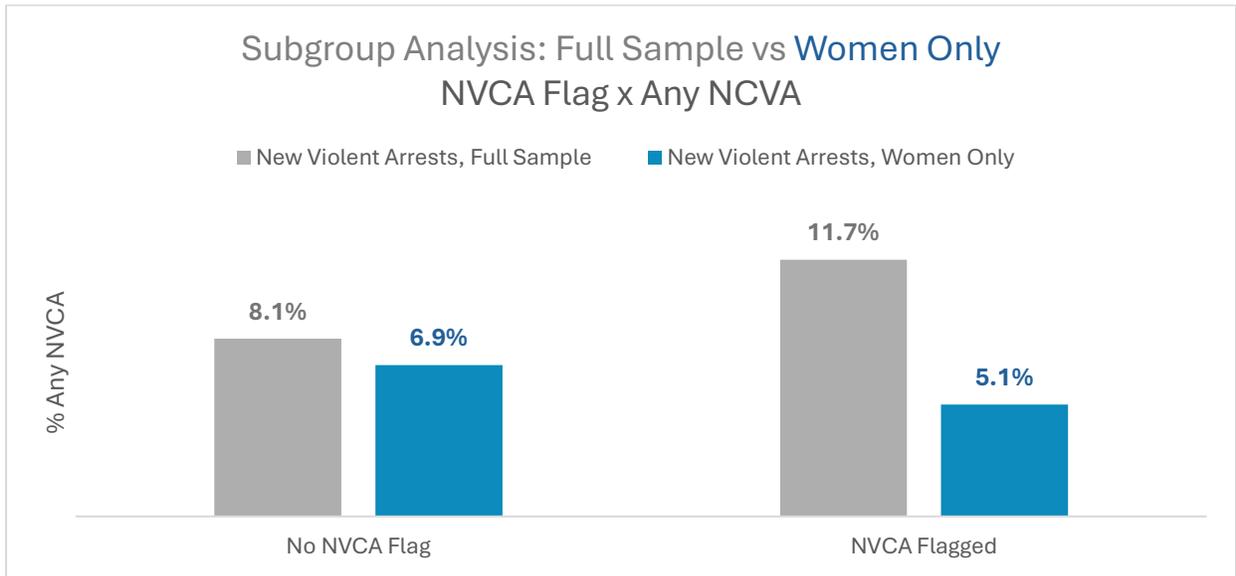
Men were more likely to have one or more new criminal arrests while released pretrial, but this was a weak statistically significant difference (χ^2 p = .003).
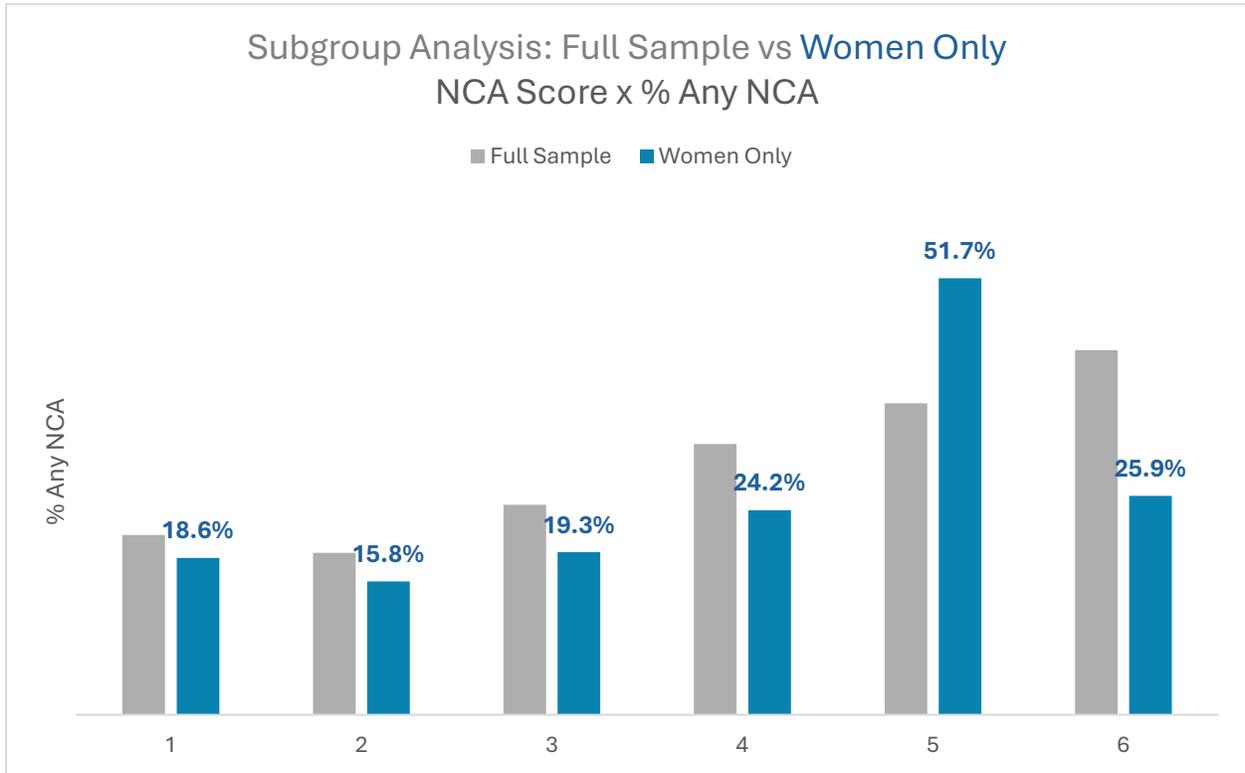
The subgroup analysis for the NCA scale for women only shows that, compared to the full sample, women have lower likelihood of any NCA at every scale score *except* score 5, where likelihood of any NCA is much higher. The Sidak post hoc analysis shows that score 5 is the *only* score to be significantly different in NCA outcomes from any other score. The effect size (η^2) for this subgroup analysis is .054, which is higher than the effect size for the whole-sample analysis (η^2 = .03). The AUC analysis produced a score of .60, which is slightly lower than the AUC for the whole-sample analysis (.61) and suggests only weak evidence of the scale's validity.

NVCA Flag and Outcomes by Sex

Subgroup Analysis: Full Sample vs Women Only
NVCA Flag x Any NCVA

■ New Violent Arrests, Full Sample   ■ New Violent Arrests, Women Only

% Any NVCA

No NVCA Flag: 8.1%, 6.9%
NVCA Flagged: 11.7%, 5.1%

The subgroup analysis revealed that women who received the NVCA flag were *less* likely to have a NVCA (5.1%) than women who were not flagged (6.9%). This is the opposite of the results from the full sample, as well as the opposite of the intention of this flag. However, *the result is not statistically significant* ($\chi^2 = 0.37$, p = 0.54). The NVCA flag may not function as intended for women.
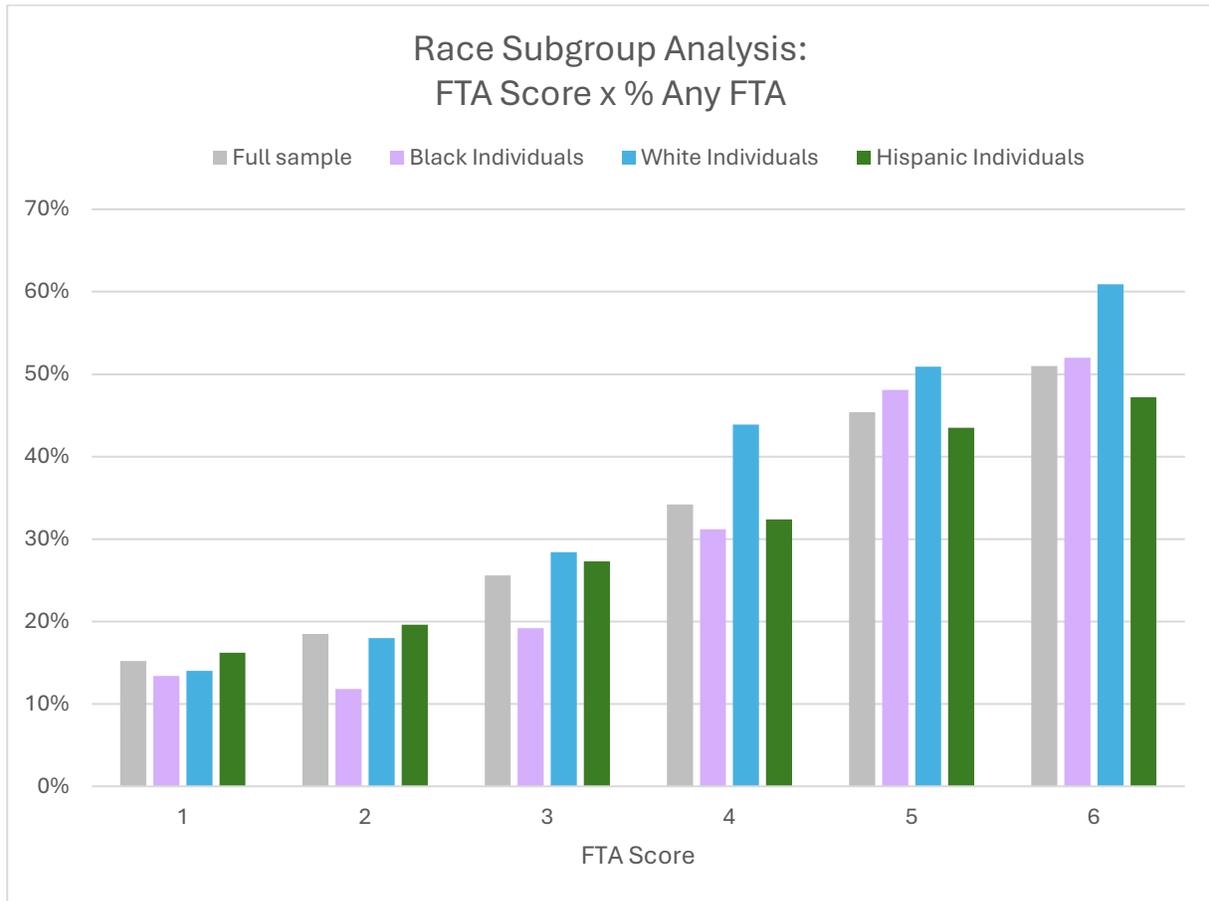
Race

The race measure in this data set includes several categories with very few cases. To avoid overgeneralizing from small samples, we have performed subgroup analyses for only the three largest groups: Black[4] individuals (n = 551), white individuals (n = 643), and Hispanic individuals (n = 2,075).

---

[4] In this document, we refer to "Black" as anyone belonging to the African diaspora. Additionally, throughout this document in line with Crenshaw (1988:1332), we capitalize "Black" as Black individuals constitute a specific cultural group and as such, require denotation as a proper noun. Those of the African diaspora have a set of shared cultures and experiences. We do not capitalize white, as white people are not a single cultural group. Crenshaw, Kimberlé (1988). "Race, Reform and Retrenchment: Transformation and Legitimation in Anti Discrimination Law." *Harvard Law Review*.

FTA Scores and Outcomes by Race

The "big picture" of the subgroup analysis shows that, overall, as the FTA score increases, the likelihood of any FTA also increases for all races. For FTA scores of 3 or more, white individuals had higher likelihoods of any FTA compared to Black and Hispanic individuals. Below, we conduct AUC analyses for each race subgroup to check for differences in the PSA's performance between groups.

FTA Scores and Any FTA Outcome

Subgroup Analysis: Proportion of Any FTA x FTA Score for Black Individuals
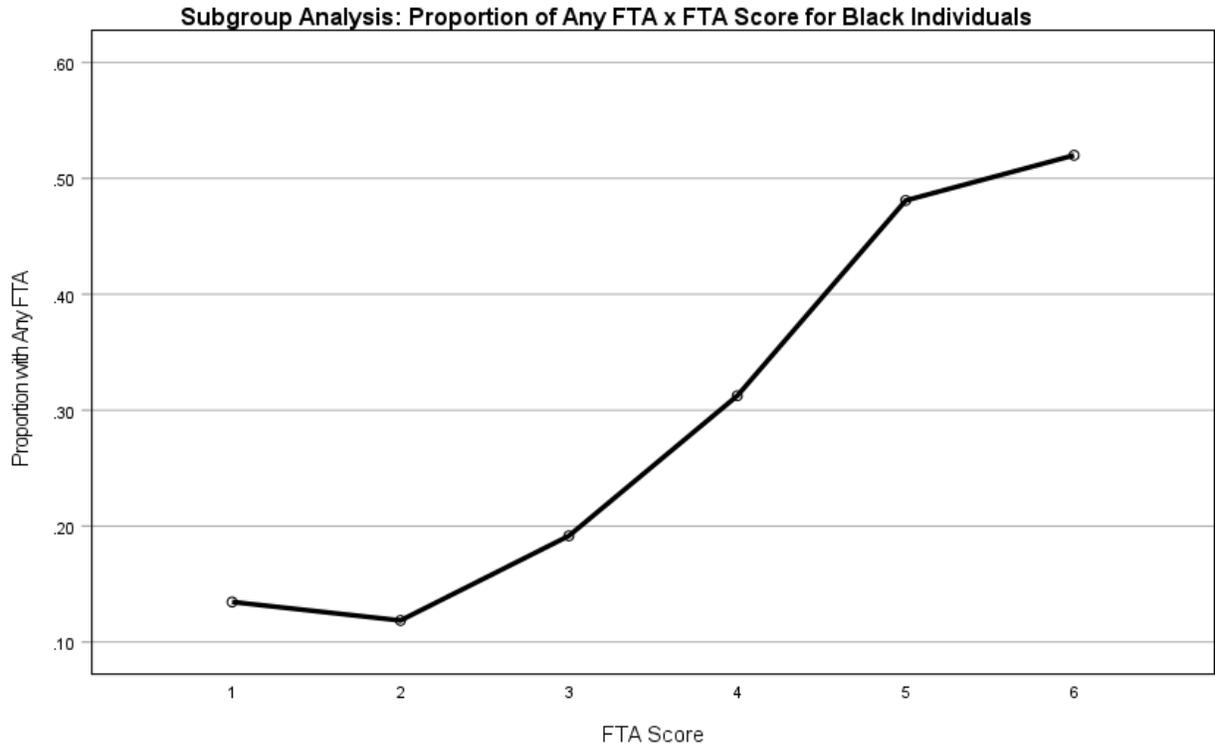
Overall, the FTA score is better at predicting FTA outcomes for Black and white individuals than for Hispanic individuals. For Black individuals, there are significant differences between high and low scale scores (F = 14.14, p < .001, η^2 = .12), with a strong AUC of .703.

Subgroup Analysis: Proportion of Any FTA x FTA Score for White Individuals
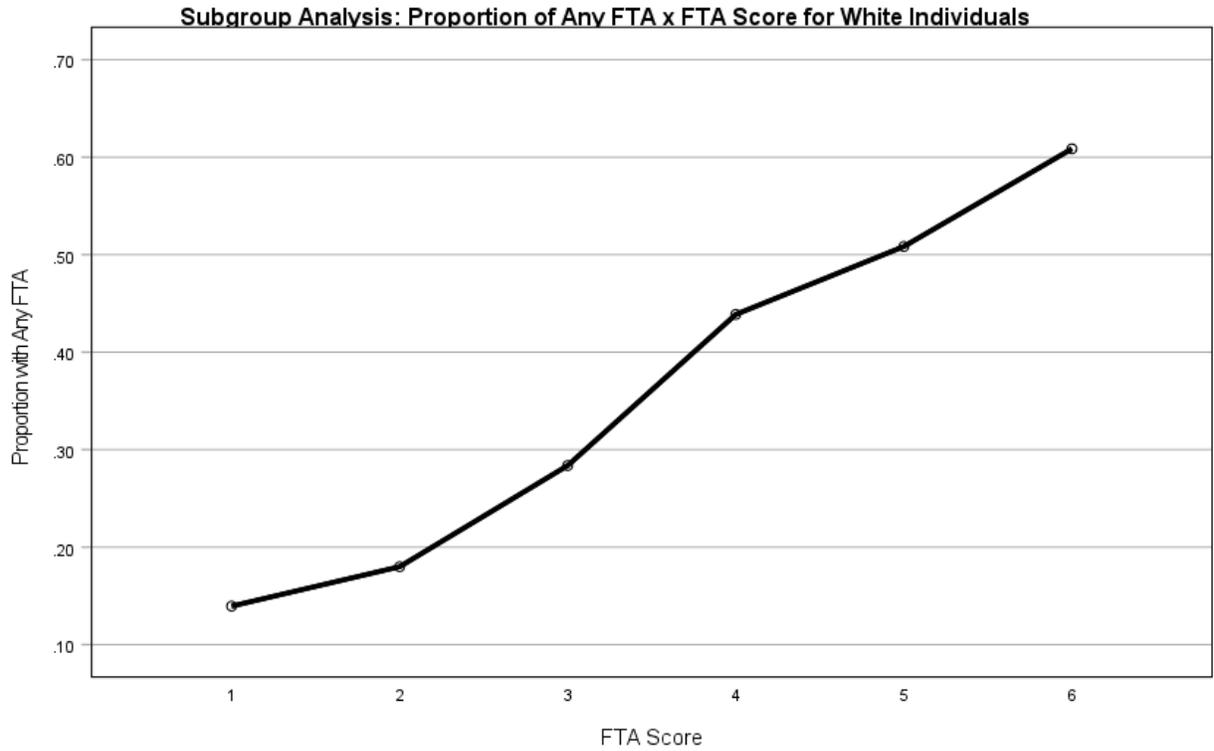
Similarly, white individuals also show significant differences in FTA outcomes across the FTA scores (F = 16.72, p < .001, η^2 = .12). The significant differences in FTA outcomes between FTA scores appear to be located mostly at the lower and higher ends of the scale – that is, outcomes differ significantly between individuals scoring 1-3 and individuals scoring 5-6, but there are fewer significant differences between closer scores.

Subgroup Analysis: Proportion of Any FTA x FTA Score for Hispanic Individuals
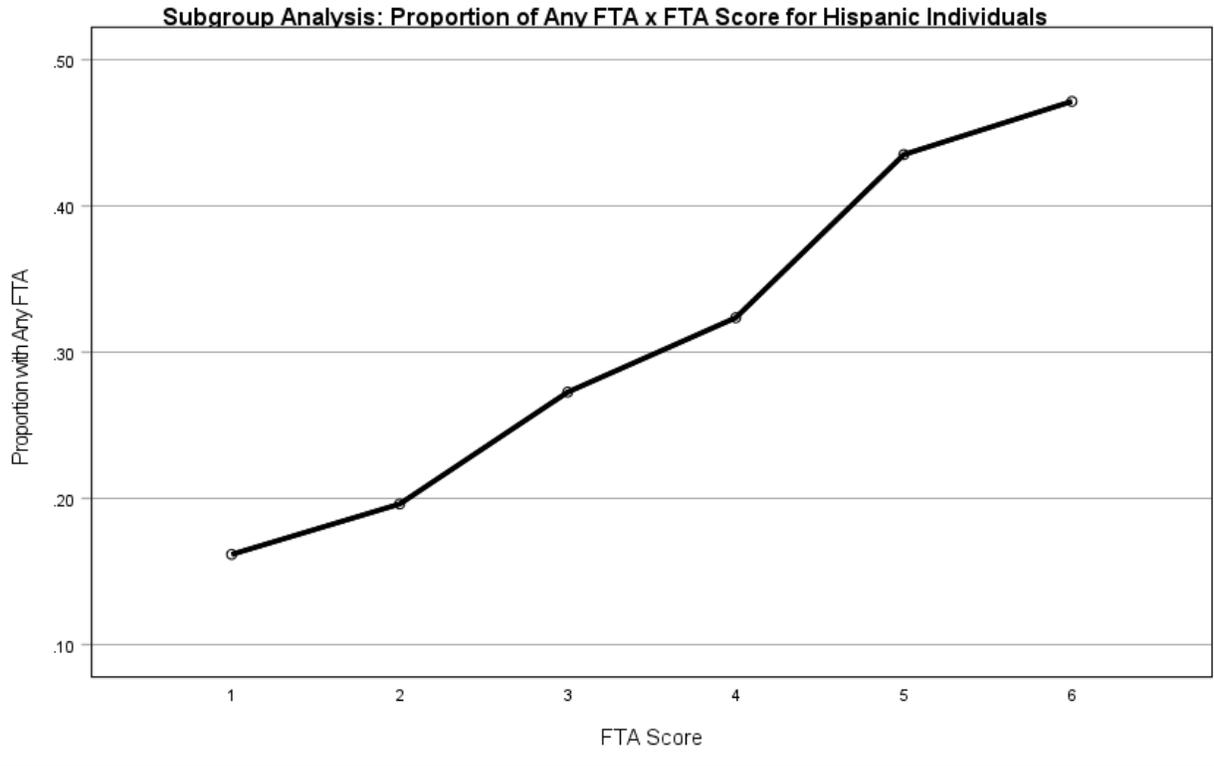
In contrast, the subgroup analysis for Hispanic individuals shows significant differences (F = 25.88, p < .001) but a smaller effect size and ROC/AUC results (η^2 = .059; AUC = .649). This is still a fairly good outcome, so it is not necessarily that the instrument underperforms for the Hispanic population but simply that it does not perform quite as well as it does for the Black and White populations.

NCA Scores and NCA Outcome by Race
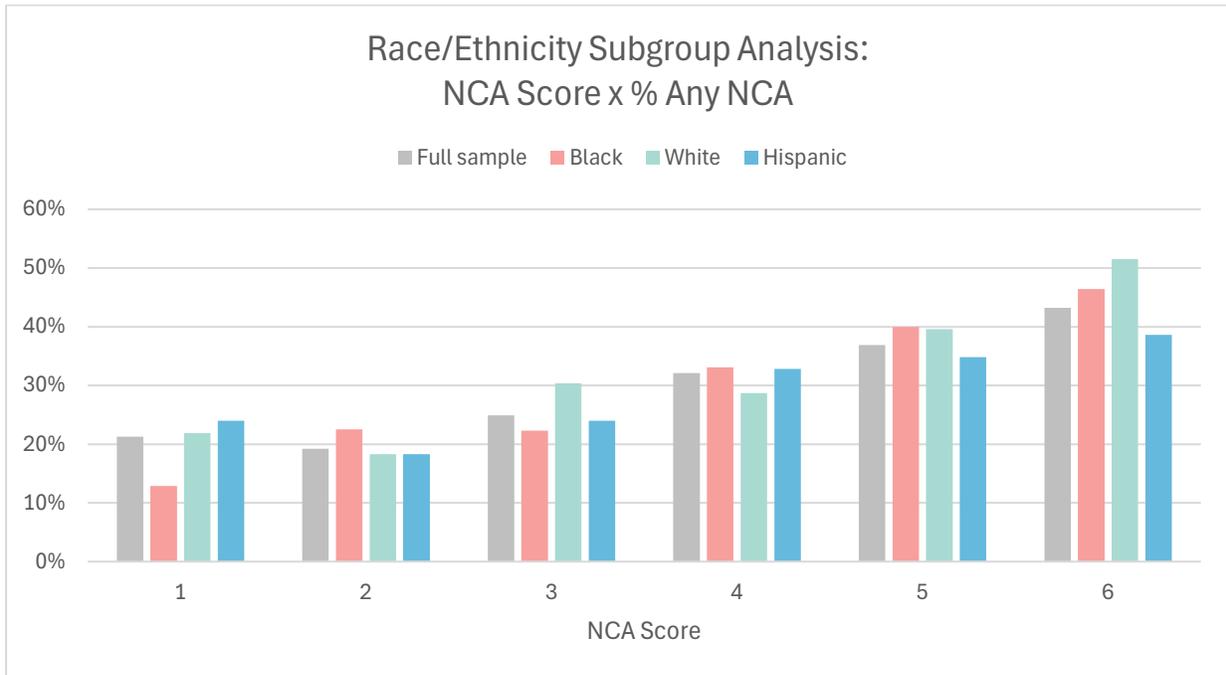
Overall, the NCA score does distinguish between NCA outcomes for Black, Hispanic, and white individuals, but again we see slightly weaker performance for Hispanic individuals. We also see smaller effect sizes for the NCA scale than for the FTA scale.

Subgroup Analysis: Proportion of Any NCA x NCA Score for Black Individuals
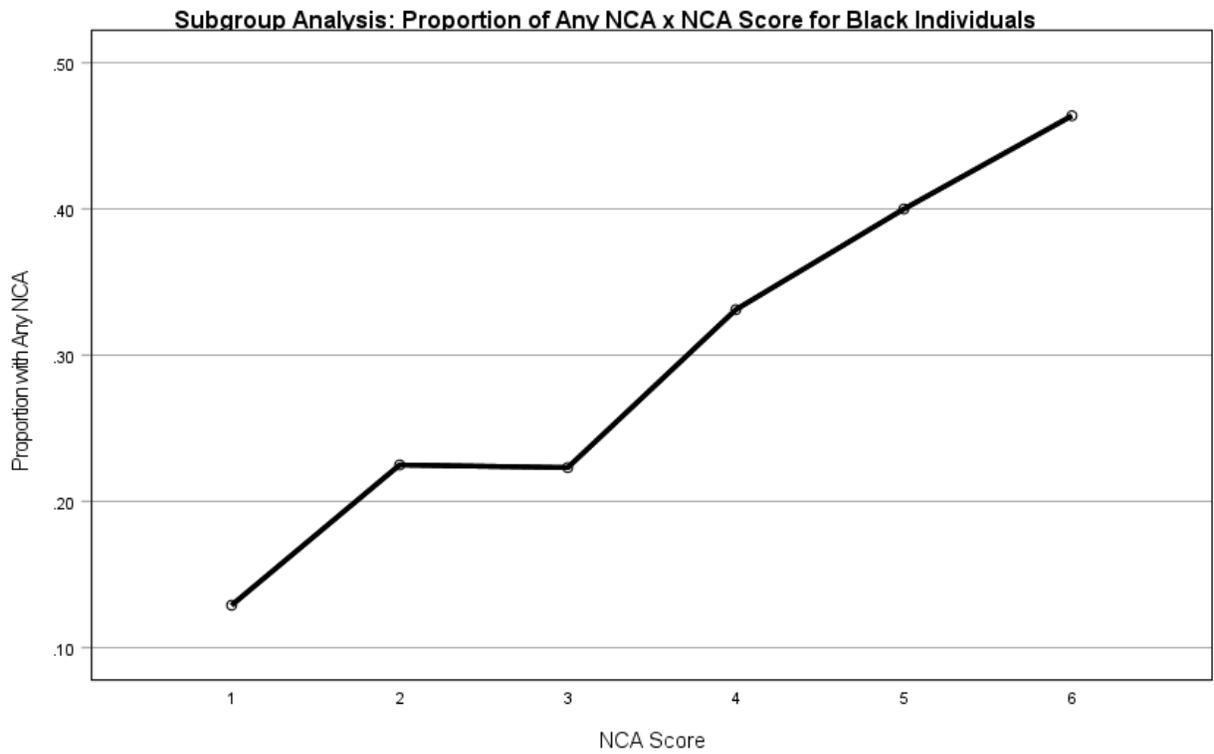
For Black individuals, there are significant differences in NCA outcomes by NCA score (F = 5.620, $p < .001$, $\eta^2 = .049$, AUC = .634), with most differences occurring between the lowest and highest scores.

Subgroup Analysis: Proportion of Any NCA x NCA Score for White Individuals
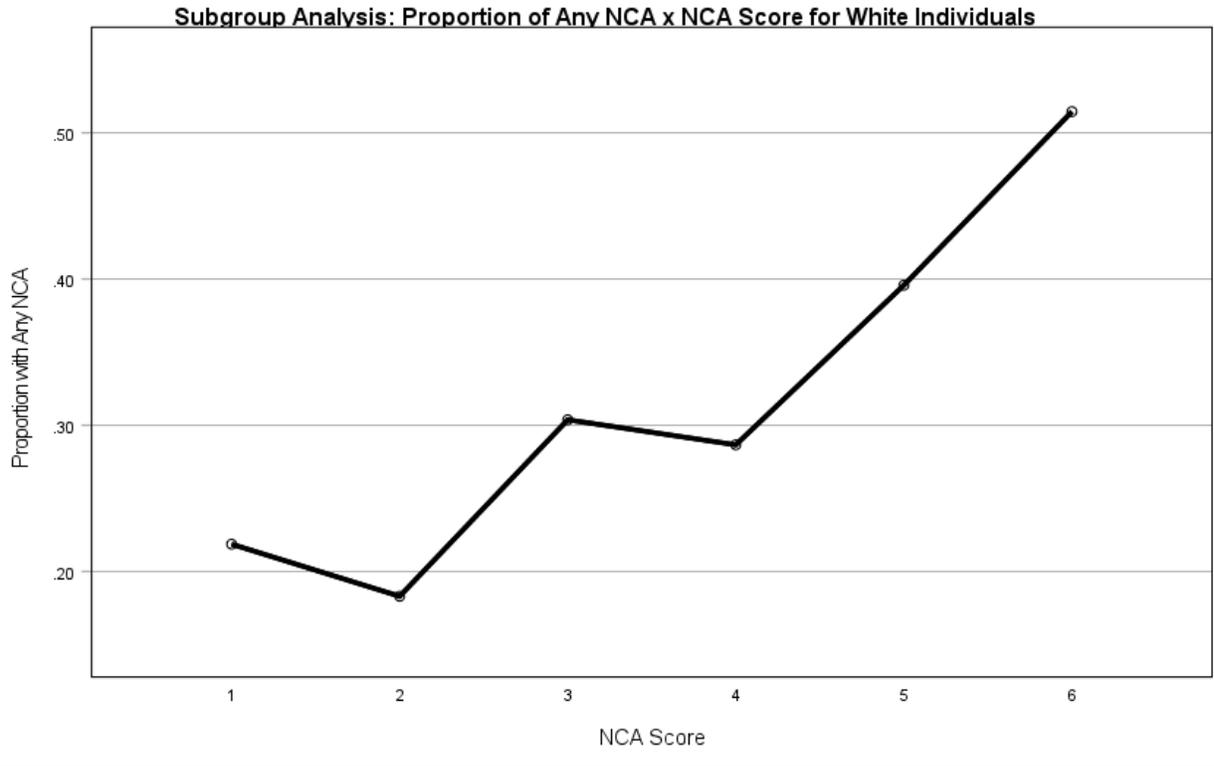
The instrument performs similarly for White individuals ($F = 6.274$, $p < .001$, $\eta^2 = .047$, AUC = .618), but the significant differences are largely between scores 2 and 6 versus the other scores.

Subgroup Analysis: Proportion of Any NCA x NCA Score for Hispanic Individuals
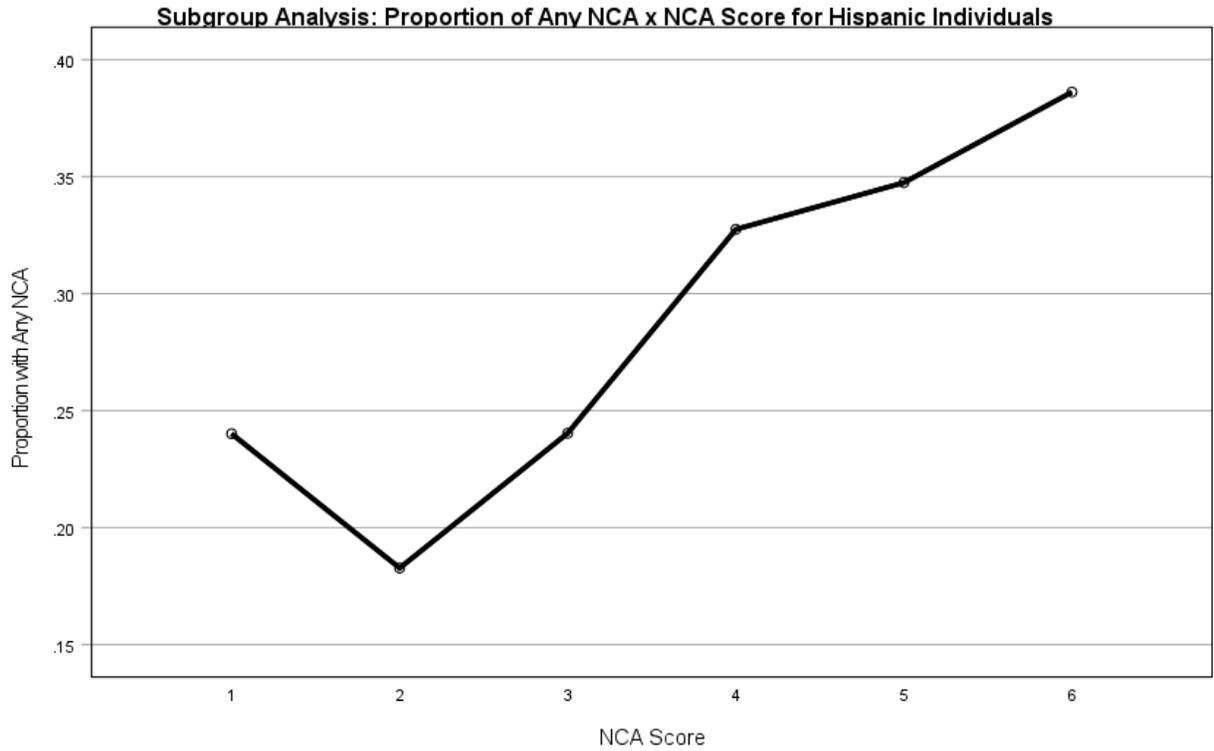
As with the FTA scale, the NCA scale also seems to be less predictive for Hispanic individuals (F = 9.867, p < .001, η^2 = .023, AUC = .588). This may be because of the unusually low proportion of any new criminal arrest for people scoring a 2 (18.3%) compared to those scoring a 1 or 3 (24%).

NVCA Flag and NVCA Outcome by Race

Race Subgroup Analysis:
NVCA Flag (Yes/No) x % Any New Violent Criminal Arrest

■ NVCA Flag "No" ■ NVCA Flag "Yes"

	Full Sample	Black Individuals	White Individuals	Hispanic Individuals
NVCA Flag "No"	8.1%	11.5%	7.0%	7.9%
NVCA Flag "Yes"	11.7%	10.7%	14.9%	11.5%

Likelihood of New Violent Criminal Arrest
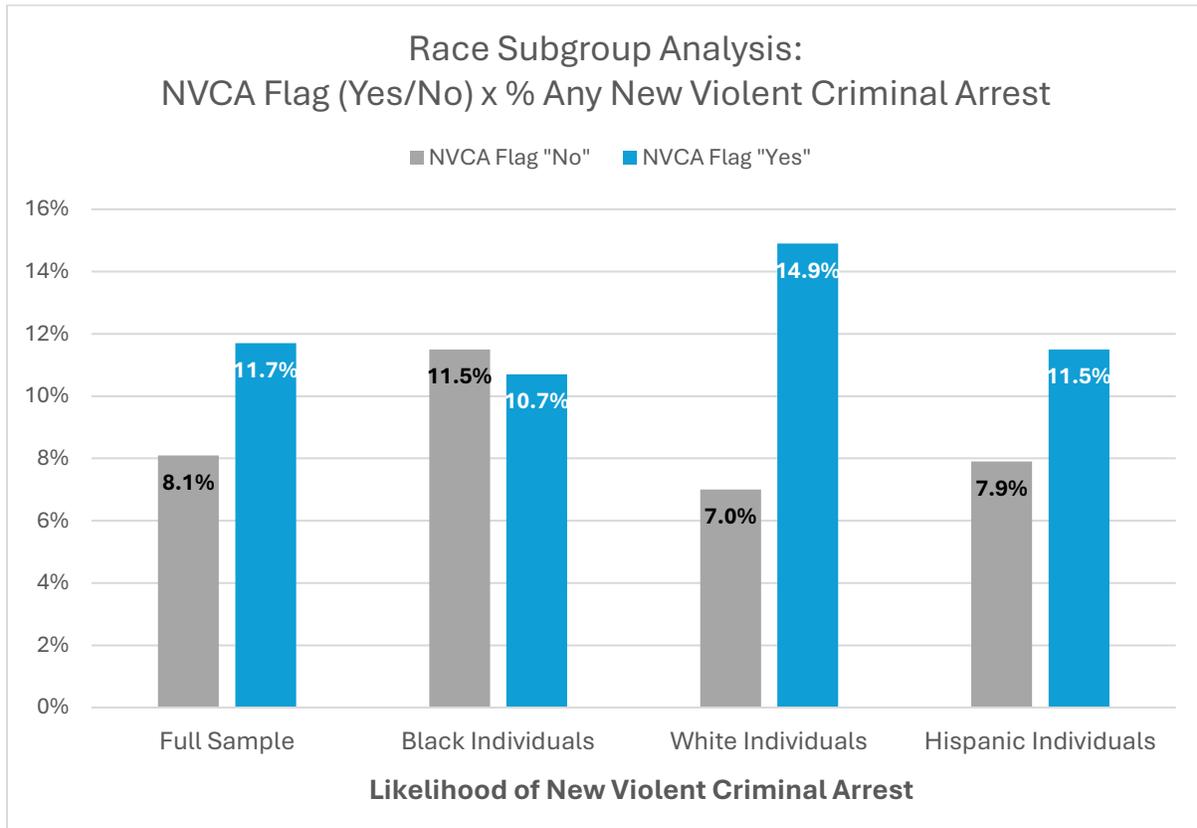
In the full sample, 8.1% of individuals flagged "No" for NVCA had a new violent criminal arrest during the observation period compared to 11.7% of those flagged "Yes." Of Black individuals flagged "No," 11.5% had a new violent criminal arrest, compared to 10.7% of those flagged "Yes" – the opposite outcome of what would be predicted, though this difference is not statistically significant (χ^2 = 0.07, p = 0.79). There is a small but significant difference in NVCA outcomes for Hispanic individuals where 7.9% of those flagged "No" had a new violent criminal arrest versus 11.5% of those flagged "Yes." This 3.6% difference is statistically significant (χ^2 = 5.25, p = 0.022). The largest statistical difference is observed for white individuals where white individuals flagged "Yes" are more than twice as likely to have a new violent criminal arrest (14.9% compared to 7% of those flagged "No"; χ^2 = 7.61, p = 0.006, Cramer's V = .11). Thus, the NVCA flag performs best for white individuals but does not perform well for Black individuals and may risk overclassifying these Black individuals into higher PSA scores.

Item Analysis for FTA, NCA, and NVCA Scores and Outcomes

FTA Items and Any FTA

Table 11, Item Analysis for FTA Scale		
	% with Any FTA	**χ² Sig., (Cramer's V)**
Q3. Pending charge at time of offense		<.001 (.171)
No	24.2%	
Yes	40.6%	
Q5a. Prior conviction		<.001 (.134)
No	24.4%	
Yes	36.7%	
Q7. Prior FTA pretrial in past 2 years		<.001 (.260)
0	20.7%	
1	34.9%	
2 or more	48.7%	
Q8. Prior FTA pretrial older than 2 years		<.001 (.170)
No	22.3%	
Yes	37.8%	

Each of the four individual items in the FTA scale was a significant predictor of any FTA. The largest effect size is seen for Q7, which measures whether the individual has "failed to appear" at a pretrial hearing in the last two years.

NCA Items and Any NCA

Table 12, Item Analysis for NCA Scale		
	% with Any NCA	**χ² Sig., (Cramer's V)**
Q1. Age at current offense		
≤ 20	23.9%	.005 (.056)
21 or 22	22.2%	
≥ 23	30.2%	
Q3. Pending charge at time of offense		<.001 (.085)
No	26.4%	
Yes	34.5%	

Q4. Prior misdemeanor conviction		<.001 (.083)
No	24.5%	
Yes	32.2%	
Q6. Prior violent conviction		**<.001 (.119)**
0	25.5%	
1	34.4%	
2	40.4%	
3	38.0%	
Q7. Prior FTA pretrial in past 2 years		**<.001 (.132)**
0	24.4%	
1	33.7%	
2 or more	38.0%	
Q9. Prior sentence to incarceration		**<.001 (.129)**
No	22.6%	
Yes	34.4%	

All items included in the NCA scale are significant predictors of any NCA outcome. The largest effect sizes, indicating the strongest predictors of new criminal arrest, are an individual's number of prior violent convictions, any "failure to appear" at a pretrial hearing in the past two years, and having a prior sentence to incarceration. Individuals who were age 23 or older were more likely to have new criminal arrests.

NVCA Items and Any NVCA

Table 13, Item Analysis for NVCA Flag		
	% with Any NVCA	**Sig., (Cramer's V)**
Q2. Current violent offense		**.010 (.044)**
No	7.7%	
Yes	10.2%	
Q2a. Current violent offense and age ≤20 years		**.017 (.049)**
No	10.0%	
Yes	13.5%	
Q3. Pending charge at time of offense		**.640 (NA)**
No	8.6%	
Yes	9.1%	

Q5a. Prior conviction			.616 (NA)
No		8.5%	
Yes		9.0%	
Q8. Prior violent conviction			.828 (NA)
No		8.6%	
Yes		9.1%	

The item analysis for the NVCA flag is helpful for understanding this flag's relatively poor performance in predicting new arrests for violent crime in this sample. Of the five included items, only two (Q2 "Current violent offense" and Q2a "Current violent offense and age ≤20 years") are significant predictors of any NVCA. The other three items, which measure other pending charges and prior convictions, are not significantly related to NVCA outcomes in this sample.

Logistic Regression Models

To run logistic regression analyses, the sex variable was coded '0' for women and '1' for men. Dummy variables were created from the race measure where a score of '1' for the variable Black means the individual is Black and 0 if the individual identifies as a race other than Black. Hispanic is '1' if the individual is Hispanic and '0' if the individual is not. Age at booking is a ratio-level measure given in years of age.

Table 14, Logistic Regression Models			
	Failure to Appear B (S. E)	**New Criminal Arrest** B (S. E)	**New Violent Criminal Arrest** B (S. E)
Constant	*-1.830 (.187)*	*-1.630 (.186)*	*-2.424 (.301)*
Man	-1.05 (.104)	.327 (.110)*	.317 (.182)
Black	-.318 (.130)*	-.094 (.125)	.294 (.193)
Hispanic	-.169 (.095)	-.182 (.094)*	.023 (.157)
Age at Booking	-.004 (.004)	-.011 (.004)*	-.015 (.006)*
NCA Score	.059 (.050)	.069 (.044)	.056 (.080)
NVCA Flag (Yes = 1)	-.163 (.103)	.216 (.050)*	.356 (.155)*
FTA Score	.343 (.044)*	-.255 (.103)*	-.002 (.069)
Model Summary (Nagelkerke R^2)	.109	.051	.015
*p < .05			

The results of the logistic regression show several notable outcomes. For the "failure to appear" outcome, we see that Black individuals are slightly less likely than white individuals to have any "failure to appear" for a pretrial hearing. We also see that, all else held equal, the FTA scale remains a significant predictor of FTA outcomes.

For the NCA scale, we see several significant predictors of new criminal arrests. Overall, men are more likely to have new criminal arrests, but Hispanic individuals are less likely to have new arrests. As age at booking increases, the likelihood of new arrests decreases. However, controlling for gender, race, age at booking, and other scale scores, the NCA score is no longer a significant predictor of new arrests. Interestingly, the likelihood of new arrests *decreases* as an individual's FTA score *increases*, suggesting that new criminal arrests and failure to appear for hearings are likely driven by different factors in individuals' lives and that failure to appear is not necessarily related to any risk to public safety.

Finally, for the outcome of any new violent criminal arrest (NVCA), as individual age increases, the likelihood of a new violent criminal arrest decreases. Holding gender, race, and other scale scores equal, the NCVA flag is predictive of new violent criminal arrests.

Key Findings

Overall, this analysis finds that the PSA performs moderately well for predicting pretrial outcomes for this sample, but there are several notable exceptions.

In some situations, the PSA struggles to distinguish between close categories at the low and high ends of the scales. That is, people at the low end (scores 1-3) usually have significantly different outcomes from people at the high end (scores 4-6), but there may not be significant differences between scores 1, 2 and 3 or between scores 4, 5, and 6. This suggests that there may not be a need for 6 different scores or that any decision-making based on these scores should be divided by 'high' and 'low' and not distinguish between, e.g., scores of 1 or 2.

This analysis found several key differences between men and women that are worthy of further investigation. For example, it may be worth doing additional research to understand why women are more likely than men to score a 5 on the FTA scale when there is no statistically significant difference in men's and women's likelihood of FTA. The FTA score considers whether there was a pending charge at the time of the offense, any prior convictions, and any prior FTA. It could be that there are systemic differences between men and women in these measures that are influencing the FTA score but not the FTA outcome.

A key takeaway from this analysis is that the New Violent Criminal Arrest (NVCA) flag is not functioning as intended, at least for specific groups within this sample. First, there is only a relatively small difference in likelihood of NVCA for those flagged "No" and those flagged "Yes" (8.1% and 11.7%, respectively). Regardless of their NVCA flag, most people (91.2%) in the sample did not have new violent criminal arrests. The subgroup analyses raise additional concerns about the function of this flag. The gender analysis shows that, among women, there is no statistically significant difference in NVCA outcomes between those with an NVCA flag and those without. In fact, in this sample the women who received an NVCA flag were *less* likely to have any NVCA while on pretrial release.

There are similar concerns about the NVCA flag's performance across different races and ethnicities. The flag performs very well for white individuals, where 7% of those flagged "No" had any NVCA versus 14.9% of those flagged "Yes." However, the NVCA flag does not predict any NVCA for Black individuals; the difference in likelihood of pretrial NVCA for Black individuals flagged "No" and "Yes" is not statistically significant. It is important to note that Black individuals in this sample were more likely to receive a "Yes" flag than the sample overall, so it is possible that Black individuals are being overclassified by this measure relevant to their actual risk of new violent arrests.

We recommend additional analyses to better understand where the PSA performs well and where it is underperforming. A larger sample of cases would allow for improved intersectional analyses of, for example, race and gender together, which would allow us to dig into the observations about women to determine if these differences are consistent for white, Black, and Hispanic women. Future analyses should also look more closely at those who were initially remanded and later released, as there may be race and gender disparities in likelihood of remand and the impact of spending some amount of time in jail instead of on pretrial release.

Appendices

Appendix A: Sample Characteristics (n = 3408)		
	Mean (St. Dev.)	*Range*
Age at Booking	34.6 (11.4)	18 - 92
Age at First Offense?	*f*	*%*
20 or younger	276	8.1
21 or 22	221	6.5
23 or older	2911	85.4
Sex[5]	*f*	*%*
Man	2820	82.7
Woman	587	17.2
Transgender[6]	1	0.0
Race of Individuals[7]		
Hispanic/Latin American	2075	60.9
White	642	18.8
Black	551	16.2
Other Asian	69	2.0
Indian	17	0.5
American Indian/Alaskan Native	16	0.5
Pacific Islander	14	0.4
Unknown	11	0.3

---

[5] We use the term "sex" to present the categorization as captured by the agency (which uses the variable name "Gender" but categories of M, F, and T; it is not clear whether this variable is based on self-reported, perceived, or assigned identity). We recognize the demographic representation of our sample may be different when considering gender identity and could include representation from individuals who identify as non-binary or gender non-conforming. We believe it is important for administrative data to consider the full spectrum of identity to understand the individual experience and disparate outcomes more adequately. This includes disaggregating the "transgender" category into transgender women, transgender men, and transgender nonbinary or gender non-conforming people.

[6] This category could not be included in the validation analysis because a single case is not sufficient for drawing conclusions about the PSA's performance for transgender individuals.

[7] We report race as the mutually exclusive categories captured by the agency. We cannot determine individuals in the sample who identify as bi- or multiracial, or who identify in other ways than what is captured by the agency. We recognize self-reported racial identity is critical for accurately reporting the true demographic profile of the sample, the individual's experience, and any disparities.

Laotian	5	0.1
Cambodian	3	0.1
Chinese	2	0.1
Other	2	0.1
Filipino	1	0.0

Appendix B: Case Characteristics and Outcomes		
	Mean (St. Dev.)	*Range*
# of FTAs	0.39 (0.69)	0 – 5
# of New Charges	0.39 (0.75)	0 - 12
Original Charge Type[8]	*f*	*%*
Felony	**2,470**	**72.5**
Drug	63	1.8
DUI	54	1.6
Domestic Violence	579	17
Other	120	3.5
Person/Animal	186	5.5
Property/ID/Fraud	461	13.5
Serious/Violent	472	13.8
Weapon	535	15.7
Misdemeanor	**938**	**27.5**
Drug	130	3.8
DUI	137	4.0
Domestic Violence	200	5.9
Other	216	6.3
Person/Animal	106	3.1
Property/ID/Fraud	112	3.3
Weapon	37	1.1
Pretrial Release Type	*f*	*%*
Granted Pretrial Monitoring	1596	46.8
Own Recognizance (OR)	434	12.7
Remains on Bond	251	7.4
Remains on Emergency Bail[9]	19	.6

[8] Based on the most serious offense only.

[9] The defendant was denied pretrial release and was remanded at arraignment but was released from custody pursuant to Emergency Bail (Zero Bail) after. The "emergency bail" condition was used by the jail during the COVID pandemic; the criteria for emergency bail release were based on agreement between the courts and the jail.

	n	%
Remains on Federal Court Order Release (FCR)	78	2.3
Remains on OR	24	.7
Remand	1006	29.5

	n	*% of remanded*
Bond After	451	44.8
Emergency Bail After	8	.8
FCR After	191	18.9
OR After	176	17.5
Pretrial After	151	15.0
Release to Program	29	2.9
Case Disposition	*f*	*%*
Dismissed	1243	36.5
Fine	1	0.0
Civil Commitment	31	0.9
Sentenced to Probation	1715	50.3
Sentenced to Jail	147	4.3
Sentenced to Local Prison	80	2.3
Sentenced to State Prison	191	5.6