JANUARY 2025

# Local Validation of the Public Safety Assessment in Pennington County, South Dakota

Policy Research Associates, Inc.

PRA
POLICY RESEARCH ASSOCIATES

## Report Prepared by:

Samantha A. Zottola, Sarah L. Desmarais, and Li Chen

## Policy Research Associates, Inc.

433 River Street, Troy, New York 12180

www.prainc.com | pra@prainc.com



## Acknowledgments

This work was supported by the John D. and Catherine T. MacArthur Foundation's Pretrial Risk Management Project. The content of this report is solely the responsibility of the authors.


## Suggested Citation

Zottola, S. A., Desmarais, S. L., & Chen, L. (2025, January). *Local validation of the Public Safety Assessment in Pennington County, South Dakota.* Policy Research Associates. https://www.prainc.com/resource-library

# Table of Contents

# Executive Summary

## Overview

In this report, we present the findings of a local validation of the Public Safety Assessment (PSA)[1] as implemented in Pennington County, South Dakota (SD). PRA began our work validating the PSA in 2023 with funding from the MacArthur Foundation's Pretrial Risk Management Project.

We examined the performance of the PSA scores in a sample of 4,570 cases that were processed in the Pennington County court, that required an initial appearance hearing, and that were PSA eligible between June 6, 2018, and December 31, 2021. Cases that are PSA eligible include people booked on new, local criminal charges and people booked following arrest on a presentence warrant. We limited our focus to cases in which the person was released from jail on a personal recognizance bond and cases that were closed at the time of data extraction. Data were drawn from the County's data management system. We examined the performance of PSA Failure to Appear (FTA) subscale scores and the New Criminal Arrest (NCA) subscale scores with respect to observed rates of failure to appear in court and new criminal arrest during the pretrial period, respectively. We also examined whether there was evidence of differences in the performance of the subscale scores (i.e., predictive bias) across subgroups defined by race (Native American and white) and gender (men and women).

## Key Findings

- About two-thirds of people who had been released from jail on personal recognizance bonds in Pennington County appeared in court and were not arrested for a new criminal charge during the pretrial period. In other words, most people were successful while awaiting their trial.
- Overall, the FTA subscale scores demonstrated poor validity in predicting failure to appear and the NCA subscale scores demonstrated fair validity in predicting new criminal arrests. These findings indicate that the PSA subscale scores do not distinguish between people at lesser and greater risk of failing to appear in court but do distinguish between people at lesser and greater risk of new criminal arrest.
- There were differences in the performance of the PSA subscale scores across groups defined by race. Both the FTA and NCA subscales were better at predicting

pretrial outcomes for white people than Native American people. Moreover, the relationship between FTA subscale scores and rates of failure to appear differed significantly for white and Native American people.

- There were also differences in the performance of the PSA subscale scores across groups defined by gender. Both subscales were consistently better at predicting pretrial outcomes for men than women. Again, the relationship between FTA subscale scores and rates of failure to appear differed significantly for men and women.

## Recommendations

Based upon the results of the current validation, we offer 5 recommendations:

1. **Limit the use of the FTA subscale scores to inform decisions about service delivery.** Our findings indicate that the FTA subscale scores have poor predictive validity and exhibit racial and gender bias, making them unsuitable for use in pretrial decision-making for the sample we analyzed. We recommend that FTA subscale scores not be used to determine pretrial release conditions. If considered, FTA subscale scores should only inform voluntary services aimed at supporting court appearances. These services should not be mandated as a condition of release, but rather offered to enhance court compliance.
2. **Consider adopting an instrument that assesses dynamic factors.** The PSA is limited to static (i.e., unchanging), largely historical items that emphasize prior criminal legal system contact. Research suggests that more proximal and dynamic items (i.e., recent, changing, or modifiable) may improve validity in predicting criminal legal outcomes among Indigenous populations. We recommend completing an assessment of dynamic factors along with the PSA to provide additional information to inform pretrial release decisions and conditions, as appropriate.
3. **Clarify override procedures and monitor their use and impact on decision-making.** At initial appearance hearings, judges in Pennington County can override PSA subscale scores. Yet, research suggests that judicial overrides can decrease predictive validity and contribute to higher rates of pretrial misconduct. We recommend evaluating how often judges use overrides, for whom, and why to help to explain the differences in the predictive validity of the PSA across race and gender. This understanding could inform strategies to support successful—and equitable—implementation of the PSA in practice.

4. **Implement continuous quality improvement (CQI) measures.** Consistent with national guidelines, we recommend the development and implementation of CQI measures, including a twice yearly case review of fidelity and consistency of the PSA scoring process, and routine booster training to refresh PSA users on administration procedures and to address issues as they arise.
5. **Conduct a validation with a more representative sample and additional data.** The current validation focused on people released on personal recognizance bonds and whose cases were closed at the time of data extraction and suffered some key limitations (e.g., no information on time at risk). We recommend another validation be conducted on a broader sample and with additional data collection to provide more comprehensive and generalizable findings regarding the performance of the PSA subscale scores in Pennington County, SD.

# Overview

## Background

Pennington County, South Dakota (SD) has been engaged in a range of pretrial reform efforts through their involvement with the MacArthur Foundation's Safety and Justice Challenge. The County's pretrial reform efforts aim to reduce the jail population overall, 90% of which represents people detained pretrial, and to reduce the overrepresentation of Native American people specifically. Though Native American people make up about 10% of the County's population, they make up about 50% of the County jail population. In 2018 Pennington County began using the Public Safety Assessment (PSA) to inform pretrial release decisions as one step toward helping them achieve their reform efforts. National guidelines recommend that risk assessment instruments are locally validated approximately every five years.[2] In this report, we describe the PSA, the purpose and process of conducting a local validation study, and the findings of our validation of the PSA in Pennington County. We conclude with recommendations based on our findings.

## Local Validation Explained

The purpose of a local validation study is to assess the performance of a pretrial risk assessment instrument as used in a particular jurisdiction. Local validation studies establish the accuracy of pretrial risk assessment results in predicting the outcomes they were designed to predict; this is known as **predictive validity**. Local validations also often (but not always) seek to determine if the assessment scores perform comparably across subgroups of people, such as those defined by race, ethnicity, or gender. There are (at least) two different strategies that are typically used to assess comparable performance. First, we can compare the accuracy of pretrial risk assessment results in predicting outcomes across groups defined by race/ethnicity or gender. This is known as **differences in predictive validity**. Second, we can assess whether the relationship between assessment results and outcomes differ across subgroups of people. This is known as testing **differential prediction**.

Validation studies should be performed in each jurisdiction that adopts a pretrial risk assessment instrument, even if that instrument has been validated in other jurisdictions.[3] This is because the validity (i.e., accuracy) with which pretrial risk assessment results forecast outcomes can be affected by local record-keeping practices; variations in

policies, statutes, and guidelines; differences in the rates of pretrial failure; differences in pretrial services; and many other local attributes. Validity can also be affected by modifications that local stakeholders make to an instrument's scoring or interpretation protocols to adjust for differences in local laws and criminal justice policies. Local validations should occur within the first several years of implementation.[4]

## The Public Safety Assessment (PSA)

The PSA is a pretrial risk assessment instrument comprised of 11 items that are coded based on information about a person's age, details of their current charge, pending charges, prior misdemeanors, prior felonies, prior violent crimes, prior failures to appear, and prior incarceration. Item codings are weighted and summed to create three separate subscales according to the scoring manual: the Failure to Appear (FTA) subscale, the New Criminal Arrest (NCA) subscale, and New Violent Criminal Arrest (NVCA) subscale. The complete list of items, response weightings, and process for creating subscale scores can be found online. The PSA is the most widely used pretrial risk assessment: hundreds of localities across at least 26 states have adopted it as part of their pretrial decision-making process.[5] Its widespread use reflects several considerations, including that it is available free to jurisdictions and was designed to be completed based on court records, without an interview.[6]

## Prior Validations of the PSA

There have been at least a dozen PSA validations. We summarize their predictive validity findings in Table 1 below. A majority have found good or better validity for all three PSA subscale scores. When studies did not find good validity, they still find the strength of prediction to be fair. Taken together, findings of prior validation studies indicate that the PSA subscale scores generally predict outcomes as intended.

**Table 1. Predictive Validity of the PSA Overall**

| *State* | *Jurisdiction* | *FTA Validity* | *NCA Validity* | *NVCA Validity* |
|---|---|---|---|---|
| **California**[7] | Los Angeles | Excellent | Excellent | Good |
| **California**[8] | San Francisco | Fair | Fair | Good |
| **Florida**[9] | Volusia County | Fair | Good | -- |
| **Georgia**[10] | Fulton County | Fair | Good | Good |
| **Illinois**[11] | McLean County | Good | Good | Fair |

| State | Jurisdiction | FTA Validity | NCA Validity | NVCA Validity |
|---|---|---|---|---|
| **Illinois**[12] | Kane County | Good | Fair | Fair |
| **Kentucky**[13] | Statewide | Good | Fair | -- |
| **Kentucky**[14] | Statewide | Good | Good | Good |
| **Ohio**[15] | Lucas County | Fair | Good | Fair |
| **Texas**[16] | Harris County | Good | Fair | -- |
| **Washington**[17] | Thurston County | Good | Good | Good |
| **Washington**[18] | Pierce County | Fair | Fair | Good |

Overall, there does not appear to be systematic bias as a function of race, but findings emphasize the importance of examining the performance of PSA subscale scores across racial and ethnic groups locally defined. A few studies have found differences in predictive validity, but when there are differences, they are not always in a consistent direction (i.e., sometimes better for white people, but other times better for people of color), and subscale scores generally (but not always) predict outcomes with at least fair accuracy across groups (see Table 2). However, there was evidence of differential prediction in relation to the FTA subscale scores, as well as the NCA subscale scores in a few jurisdictions. In other words, the relationships between the FTA and NCA subscale scores and the outcomes failure to appear and new criminal arrest function differently across race in some instances (see Table 3).

### Table 2. Differences in Predictive Validity of PSA Subscale Scores by Race

| | | FTA Validity | | NCA Validity | | NVCA Validity | |
|---|---|---|---|---|---|---|---|
| **State** | **Jurisdiction** | **White** | **People of Color** | **White** | **People of Color** | **White** | **People of Color** |
| **California**[19] | San Francisco | Good | Fair | Fair* | Fair* | Good* | Good-Fair* |
| **Kentucky**[20] | Statewide | Good* | Fair* | Good | Good | Good | Good |
| **Washington**[21] | Thurston County | Good | Good | Good | Excellent | Good | Good |
| **Washington**[22] | Pierce County | Poor* | Fair* | Good | Good | Good | Good |

*Note.* *Indicates a significant difference in the AUC values.

**Table 3. Differential Prediction of PSA Subscale Scores by Race**

| *State* | *Jurisdiction* | *FTA Validity* | *NCA Validity* | *NVCA Validity* |
|---|---|---|---|---|
| **Georgia**[23] | Fulton County | Yes | Yes | No |
| **Illinois**[24] | McLean County | Yes | No | No |
| **Illinois**[25] | Kane County | Yes | Yes | No |
| **Kentucky**[26] | Statewide | Yes | No | No |
| **Ohio**[27] | Lucas County | Yes | No | No |
| **Texas**[28] | Harris County | No | Yes | No |
| **Washington**[29] | Thurston County | No | No | No |
| **Washington**[30] | Pierce County | No | No | No |

Overall, there does not appear to be systematic bias as a function of gender, but again, findings emphasize the importance of local validation, as evaluations have sometimes found better predictions for women and other times found better predictions for men. However, when there are differences, PSA subscale scores generally (but not always) predict outcomes with at least fair accuracy for both men and women (see Table 4). That said, there was consistent evidence of differential prediction for the FTA subscale specifically. In other words, the relationship between FTA subscale scores and the outcome failure to appear functions differently for women and men, though the difference is not always in the same direction (see Table 5).

**Table 4. Differences in Predictive Validity of PSA Subscale Scores by Gender**

| | | *FTA Validity* | | *NCA Validity* | | *NVCA Validity* | |
|---|---|---|---|---|---|---|---|
| *State* | *Jurisdiction* | *Women* | *Men* | *Women* | *Men* | *Women* | *Men* |
| **California**[31] | San Francisco | Good | Fair | Fair* | Fair* | Good* | Good-Fair* |
| **Kentucky**[32] | Statewide | Good* | Fair* | Good | Good | Good | Good |
| **Washington**[33] | Thurston County | Good | Good | Good | Excellent | Good | Good |
| **Washington**[34] | Pierce County | Poor* | Fair* | Good | Good | Good | Good |

*Note.* *Indicates a significant difference in the AUC values.

**Table 5. Differential Prediction of PSA Subscale Scores by Gender**

| *State* | *Jurisdiction* | *FTA Validity* | *NCA Validity* | *NVCA Validity* |
|---|---|---|---|---|
| **Georgia**[35] | Fulton County | Yes | No | No |
| **Illinois**[36] | McLean County | -- | -- | -- |
| **Illinois**[37] | Kane County | -- | -- | -- |
| **Kentucky**[38] | Statewide | -- | -- | -- |
| **Ohio**[39] | Lucas County | Yes | No | No |
| **Texas**[40] | Harris County | -- | -- | -- |
| **Washington**[41] | Thurston County | Yes | No | No |
| **Washington**[42] | Pierce County | Yes | No | No |

## Pennington County Local Validation

In Pennington County, booking or pretrial staff complete the PSA using both state and national criminal records to score each item. Booking staff complete the PSA shortly after a person is booked into the jail if they are eligible to be released directly from jail per the County's release guidelines. If a person is not eligible for release directly from jail and must be held until first appearance (hold reasons include open warrants and arrests on probation, parole, or bond condition violations), pretrial staff complete the PSA before the first appearance hearing. Completed PSAs are entered in an electronic management system and can be accessed by booking staff, pretrial staff, and members of the judiciary. PSA subscale scores are also provided to the State's Attorney's Office and the Public Defender's Office before the first appearance hearing. PSA subscale scores are then used to inform release decisions made by jail staff for people who are released directly from the jail, and by judges, for people who are held in jail until their first appearance.

In this report, we describe a local validation of the PSA subscale scores in Pennington County with respect to rates of court appearance (i.e., failure to appear) and new criminal arrest during the pretrial period. We assessed the validity of the PRA results in predicting these two outcomes overall and across subgroups of people defined by race and gender. We do not report on the PSA's validity in predicting new violent criminal arrests during the pretrial period because this outcome was not recorded consistently in the data.

## Data Source

We analyzed data maintained by Pennington County. Specifically, Liz Hassett, who was serving as the Pennington County Grant Manager at the time of this validation, pulled the data directly from the County's data management system. Hassett verified and deidentified the data then shared the file with PRA research staff for data cleaning and analysis.

## Sample

The full dataset shared with us included all cases (*N* = 20,326) processed in the Pennington County Court that required an initial appearance hearing and were PSA-eligible between June 6, 2018, and December 31, 2021. PSA-eligible cases include people booked on new, local criminal charges and people booked following arrest on a pre-sentence warrant. Cases that are not PSA-eligible include people booked on violations of bond, probation, or parole conditions; people booked for Immigration and Customs Enforcement (ICE) holds; and people booked on holds for other counties.

To create our analytic sample, we excluded cases that were still open (*n* = 8,320) at the time of data extraction, cases for which judges dismissed (*n* = 1,365) or declined to pursue (*n* = 377) the charges, as well as cases in which a person pled guilty at first appearance, thus resolving their case (*n* = 1,020). We also excluded cases in which the person charged was assigned bail (*n* = 2,956) because we had no way of knowing whether they were able to pay their bail and be released or not); cases in which the person was held without bail (*n* = 118) because they would not have been at risk of failure to appear or new criminal arrest; cases in which the person was released on probation (*n* = 32) because they are not within the pretrial period; and cases for which a person's bond information was missing (*n* = 1,568) because we could not determine if they were in the community at any point during the pretrial period.


Our final analytic sample included 4,570 cases for people who were released on a personal recognizance bond and thus at risk of the two primary outcomes of interest: failure to appear and arrest for a new crime during the pretrial period. Our sample may include people who were booked more than once during the study period; the person-level identification number was removed as part of the data de-identification process. However, analyzing data at the case level is standard practice for the evaluation of pretrial risk assessment results and PSA scores, specifically.[43,44]

## Variables

Conducting a validation study of the PSA assessments completed in Pennington County required the sets of variables described below.

### PSA Subscale Scores

The PSA results analyzed in this validation include the subscale scores assigned to each case for two PSA subscales: *the FTA subscale* and *the NCA subscale*. In line with standard practice in the field, we use the subscale scores, not the raw scores.

### Outcome Variables

The validation examined two outcomes of interest: 1) *failure to appear*, and 2) *new criminal arrest*. Failure to appear was a binary indicator of whether a person did or did not fail to appear for any court dates that were scheduled throughout the duration of their case. New criminal arrest was a binary measure of whether a person was or was not arrested for a new criminal charge during the pretrial period. Staff record a new criminal arrest when a person encounters law enforcement and is brought to the jail. Staff record only the first instance of failure to appear or new criminal arrest; subsequent instances of either outcome are not recorded. Outcomes are mutually exclusive because the system does not allow staff to record a person as experiencing both outcomes. Thus, people are recorded as having failed to appear, as having a new criminal arrest, or as being successful (i.e., no failure to appear or rearrest) during the pretrial period.

### Demographic Characteristics

The demographic characteristics included the *race/ethnicity* and *gender* of each person charged with an offense. Race/ethnicity categories recorded in the jail records included: Native American, white, Hispanic, and Asian or Pacific Islander. We limit our analyses to comparisons of Native American and white people because these groups were the only two racial/ethnic groups with a sufficient number of people to support statistical analyses. Gender was limited to men and women as these are the only categories recorded by the jail.

## Analysis

In this report, we present three types of findings, each representing a critical step in the process of validating a risk assessment instrument (pretrial or otherwise).

### Step 1

The first step is to **describe** sample characteristics, pretrial outcomes, and PSA results. To do so, we provide descriptive statistics (which includes averages, counts, and percentages) and bivariate comparisons (which include testing whether the PSA risk levels differ between groups in statistically significant ways). These statistics provide necessary information but are not sufficient alone for conducting a validation study. The information is useful for county partners to understand how PSA results look overall and across groups. There are no cutoffs or benchmarks for these findings. Therefore, we cannot say whether the distribution or difference in PSA subscale scores or rates of pretrial outcomes overall or between two groups is acceptable or not acceptable.

### Step 2

The second step is to present measures of **predictive validity** that speak to the accuracy with which PSA results forecast pretrial outcomes. Specifically, we measured predictive validity using a performance indicator called an area under the curve (AUCs) of receiver operating characteristic curves. In a validation study, an AUC represents the likelihood that a randomly selected person who failed to appear in court or received a new criminal arrest during their pretrial period received a higher risk score than a randomly selected person who appeared in court or was not a recipient of a new criminal arrest during their pretrial period. The values range from 0 to 1, where .50 represents chance levels of prediction. Values above .50 indicate that increases in the risk score are associated with increases in the likelihood of pretrial failure. In other words, values above .50 indicate that predictive validity is better than flipping a coin. The closer the value is to 1.00, the better the predictive validity. Values below .50 indicate that decreases in the risk score are associated with increases in the likelihood of pretrial failure (or, alternatively, that increases in the risk score are associated with decreases in the likelihood of pretrial failure).

We provide the guidelines we used for interpreting the practical significance of AUC values in Table 6 below, based upon benchmarks[45] used in the field of risk assessment.[1]

**Table 6. Benchmarks for Interpreting the Predictive Validity**

| *Predictive Validity* | *AUC Values* |
|---|---|
| **Poor** | < .55 |
| **Fair** | .55 – .63 |
| **Good** | .64 – .70 |
| **Excellent** | .71 – 1.00 |

AUC values can be used to compare differences in predictive validity between subgroups of individuals (e.g., between men and women or between Native American people and white people). However, we do not infer bias, including gender or racial bias, based upon differences in AUC values. Instead, decisions regarding how much of a difference in predictive validity between groups is acceptable are not statistical in nature. Community partners should examine the level of predictive validity indicated by AUC values within each subgroup and consider: 1) whether they are acceptable within each subgroup; and 2) whether there are meaningful differences in practical significance between groups (e.g., good performance in one subgroup but poor performance in another). Even if there are statistically significant differences in AUC values between groups, this is not necessarily problematic if the overall level of predictive validity is acceptable within each subgroup. However, if predictive validity within a subgroup is lower than deemed acceptable and/or is meaningfully different between groups, community partners may wish to take corrective action, such as changing the guidelines for applying the risk assessment results to inform decision-making, making changes to the scoring system, or other strategies.

### Step 3

The third step is to examine **differential prediction**, that is, whether PSA results predict pretrial outcomes in different ways across groups defined by race and gender. We examined differential prediction by using logistic regression analyses to statistically determine whether the average PSA subscale score related to the average rate of pretrial outcomes in the same way for each subgroup. These analyses modeled the prediction of pretrial outcomes by the PSA subscale scores, by the grouping variable (e.g., gender or race), and by a term that reflects the interaction of the PSA subscale scores with the group variable (e.g., PSA subscale scores x gender, or PSA subscale scores x race). These interaction terms allow us to measure and statistically estimate the influence of the grouping variable on the prediction of the PSA subscale score. If that interaction term is a statistically significant predictor of pretrial outcomes, then there is evidence that the average PSA subscale score relates to the average rate of pretrial outcomes differently across groups, and thus, there is evidence of differential prediction (i.e., bias).

# PSA Validation Findings

## Step 1: Descriptive Results

### Case Characteristics

The sample used to validate the PSA included 4,570 cases in which people were released on a personal recognizance bond between June 6, 2018 and December 31, 2021. The average age across cases was 33.47 (*SD* = 11.10). Almost two-thirds of cases in our analytic sample were for Native American people (62.9%, *n* = 2,876) and about one-third were for white people (31.8%, *n* = 1,454). Black people (3.0%, *n* = 138), Hispanic people (1.4%, *n* = 63), and Asian or Pacific Islander people (0.7%, *n* = 35) represented very small proportions of our analytic sample. Two-thirds of cases were for men (66.4%, *n* = 3,015) and a third of cases were for women (33.6%, *n* = 1,523); gender data were missing for 0.7% (*n* = 32) of cases. There was an average of 2.20 (*SD* = 1.48) charges associated with each case. For just over half of the cases, the highest charge was a misdemeanor (56.0%, *n* = 2,558). The most common charges in our analytic sample were: driving under the influence, possession of a controlled drug or substance schedule two, possession or use of drug paraphernalia, impersonation to deceive law enforcement, and simple assault.

### Pretrial Outcomes

Overall, people failed to appear for a court date in just over one-third of cases (37.2%, *n* = 1,700). Failure to appear rates were significantly higher among cases with Native American people (42.2%, *n* = 1,215) compared to white people (27.2%, *n* = 403), but still represented a minority of cases, $x^2(1) = 87.11$, $p < .001$. Rates of failure to appear were comparable for men (36.4%, *n* = 1,096) and women (38.7%, *n* = 590), $x^2(1) = 2.47$, $p = .116$.

Overall, people were rearrested during the pretrial period in just over one-third of the cases included in our analytic sample (37.1%, *n* = 1,697). Pretrial rearrest rates were significantly higher for Native American people (40.3%, *n* = 1,160) compared to white people (30.9%, *n* = 450), $x^2(1) = 36.42$, $p < .001$. As with for failure to appear, pretrial rearrest rates were comparable for men (37.7%, *n* = 1,136) and women (36.3%, *n* = 553), $x^2(1) = 0.81$, $p = .368$.

## PSA Scores

### FTA Subscale Scores

In our overall analytic sample, FTA subscale scores represent the full range of scores (from 1-6) with an average of 3.75 (*SD* = 1.63). Figure 1 displays the distribution of cases across FTA subscale scores. The most common FTA subscale score was 5, though scores of 3 – 6 were received in roughly the same number of cases. Only about 1 in 10 cases (11.2%, *n* = 510) received the lowest subscale score 1. This distribution is counter to expectations. Specifically, we typically expect to find lower subscale scores among people who are released on personal recognizance bonds as this form of pretrial release is often used for people who are (perceived to be) lower risk for pretrial misconduct, and thus, are not seen as requiring conditions or other strategies to support their success during the pretrial period.

**Figure 1. Distribution of Cases Across FTA Subscale Scores**

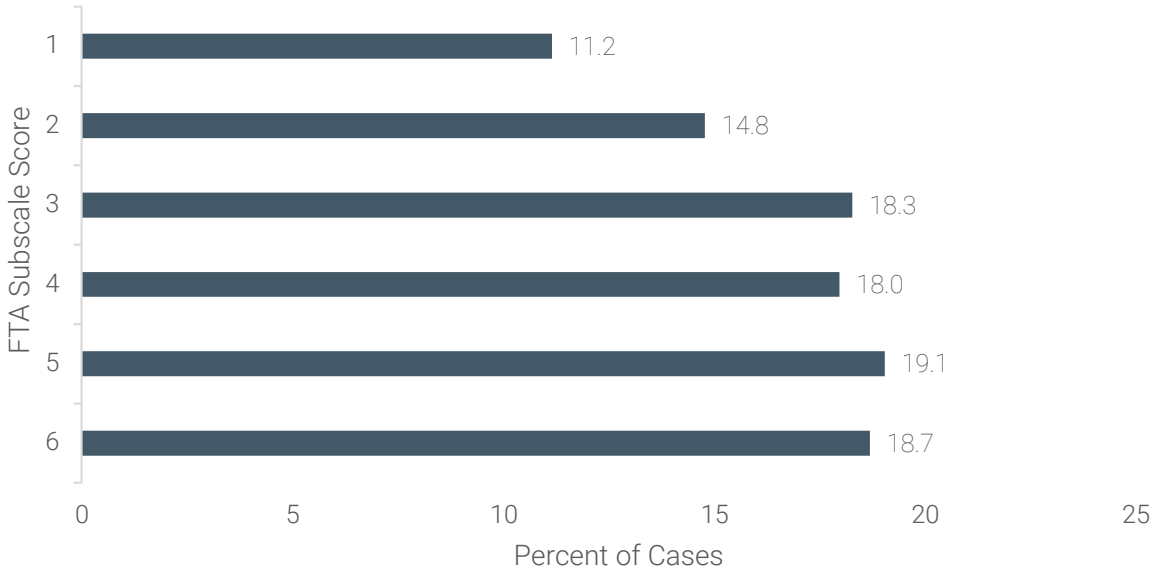



Figure 2 shows the distribution of cases across FTA subscale scores for Native American and white people separately. The most common FTA subscale score was much higher for Native American people (i.e., 6, which was received in 24.1% of cases, *n* = 692) than for white people (i.e., 2, which was received in 22.5% of cases, *n* = 327). This pattern held across the individual FTA subscale scores. Accordingly, FTA subscale scores were significantly higher for cases involving Native American people, with an average score of 4.11 (*SD* = 1.55), compared to cases involving white people, with an average score of 3.10 (*SD* = 1.58), $t(4,328) = -20.23$, $p < .001$, $d = 1.56$.

## Figure 2. Distribution of Cases Across FTA Subscale Scores for Native American and White People

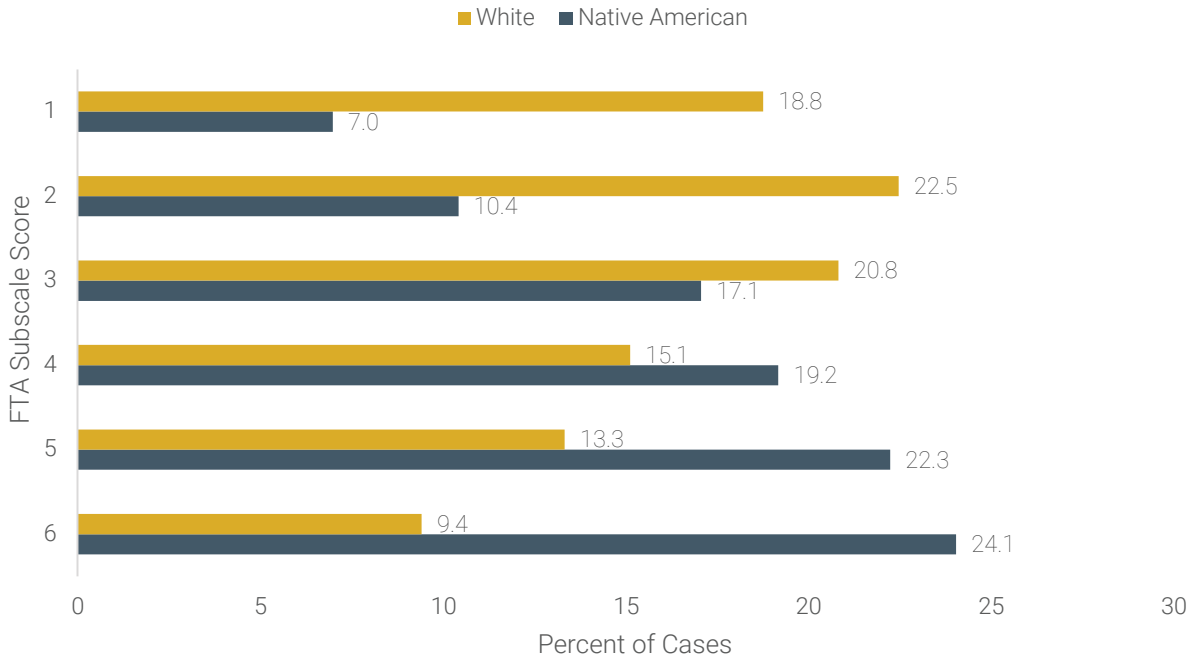
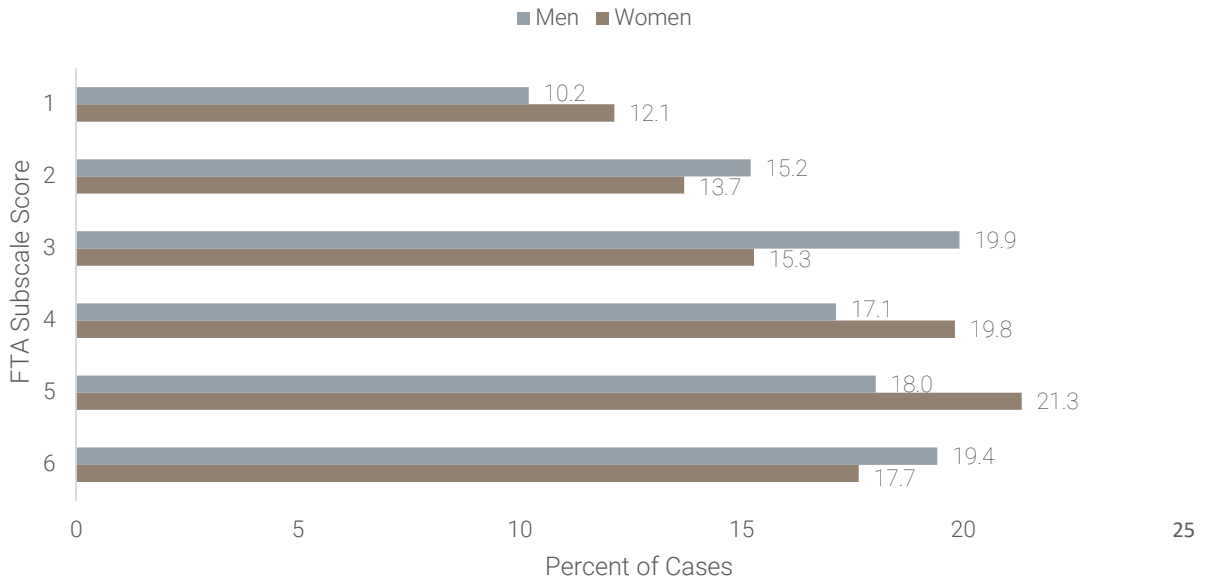



Figure 3 shows the distribution of cases across FTA subscale scores for men and women separately. The most common FTA subscale score was lower for men (i.e., 3, which was received in 19.9% of cases, $n$ = 601) than for women (i.e., 5, which was received in 21.3%, $n$ = 325). However, FTA subscale scores were very comparable overall: cases involving men had an average FTA subscale score of 3.76 ($SD$ = 1.62) and cases involving women had an average FTA subscale score of 3.77 ($SD$ = 1.63), $t(4{,}536) = -0.31$, $p = .755$, $d = 1.63$.

**Figure 3. Distribution of Cases Across FTA Subscale Scores for Men and Women**




## NCA Subscale Scores

Figure 4 shows the distribution of cases across NCA subscale scores in our overall analytic sample. As for the FTA subscale scores, the NCA subscale scores represent the full range of scores (from 1-6) with an average of 3.80 (*SD* = 1.53). The most common NCA subscale score was 4, which was received in close to one-quarter of cases (24.0%, *n* = 1,096) and the least common NCA subscale score was 1, which was received in fewer than 1 in 10 cases (7.7%, *n* = 353) As we discussed in relation to the FTA subscale scores, the distribution of NCA subscale scores is counter to expectations as, often, lower subscale scores are more common than higher subscale scores among people who are released on personal recognizance bonds and thus, are not seen as requiring conditions or other strategies to support their success during the pretrial period.

**Figure 4. Distribution of NCA Subscale Scores**

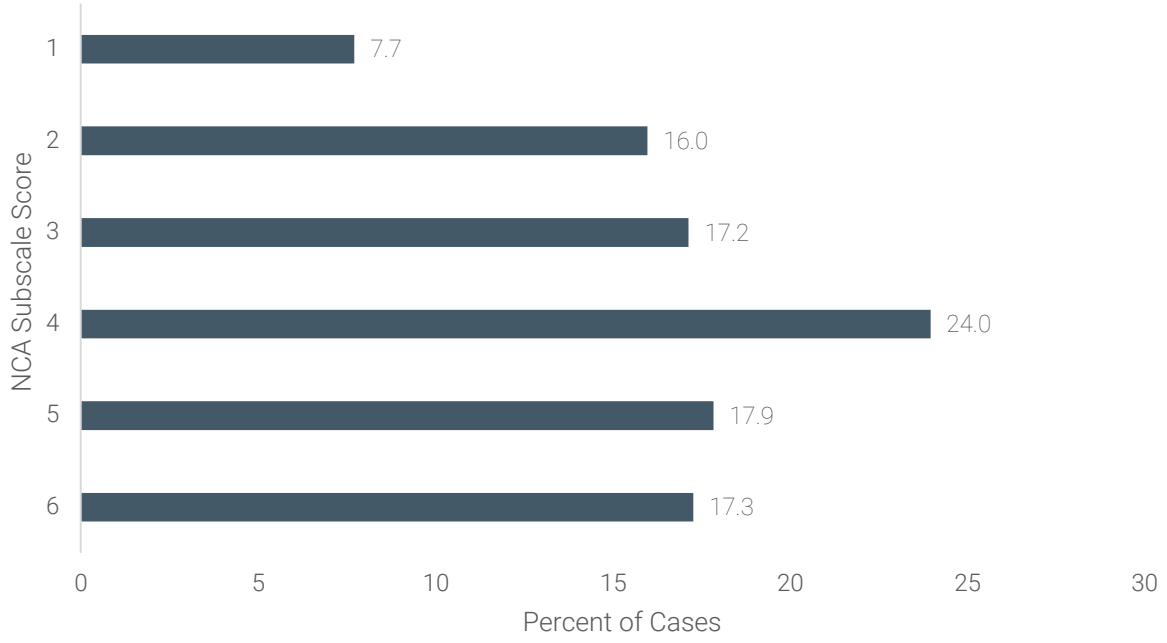



Figure 5 shows the distribution of cases across NCA subscale scores for Native American and white people separately. The most common NCA subscale score was higher for Native American people (i.e., 4, which was received in 27.0% of cases, $n$ = 777) than for white people (i.e., 2, which was received in 23.9% of cases, $n$ = 348). This pattern of seeing a greater proportion of higher scores among Native American people and a greater proportion of higher lower scores among white people held across the individual NCA subscale scores. Accordingly, NCA subscale scores were significantly higher for cases involving Native American people, with an average score of 4.11 ($SD$ = 1.43), compared to cases involving white people, with an average score of 3.24 ($SD$ = 1.55), $t(4,328) = -18.40$, $p < .001$, $d = 1.57$.

## Figure 5. Distribution of NCA Subscale Scores by Race

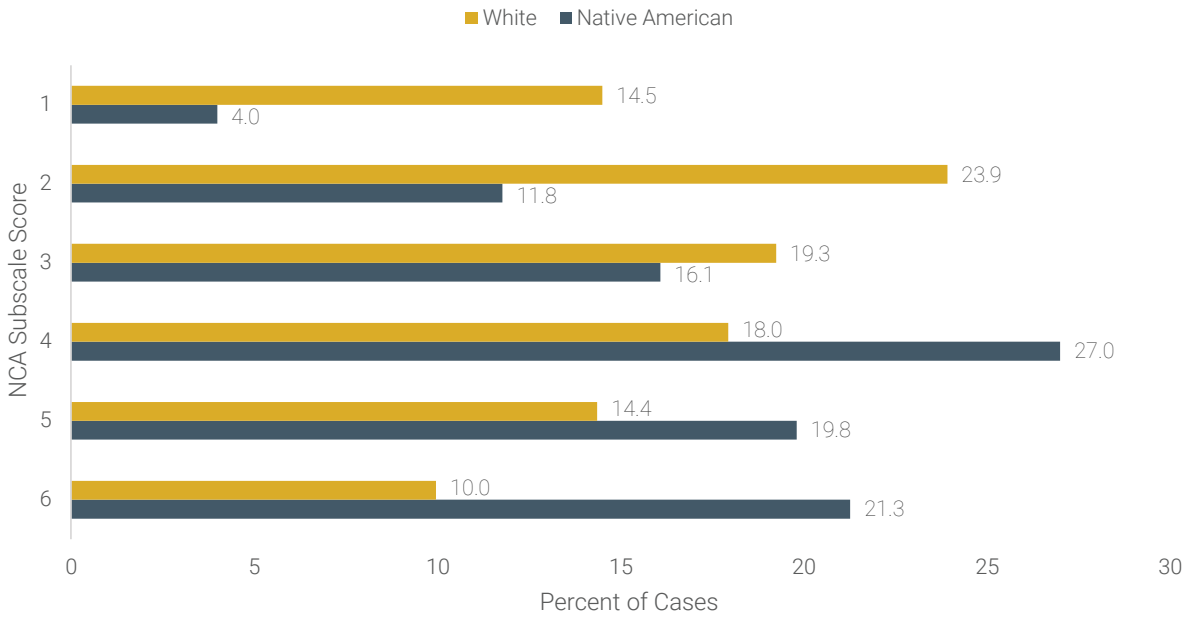
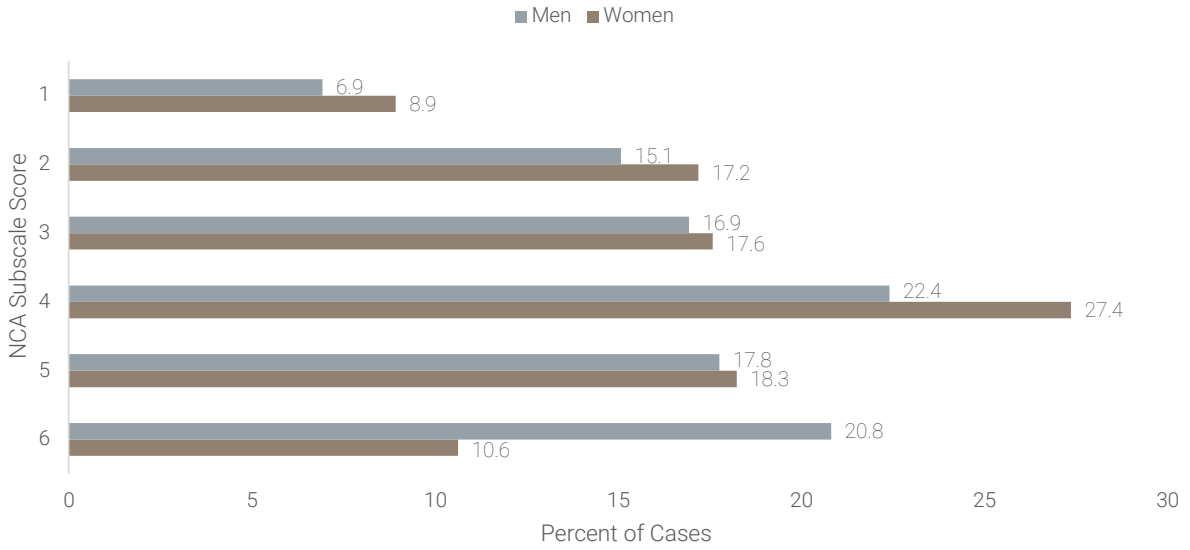



Figure 6 shows the distribution of NCA subscale scores for cases involving men and women separately. Generally speaking, NCA subscale scores were quite comparable with a few exceptions. The most common NCA subscale score was 4 for both men (22.4%, $n$ = 676) and women (27.4%, $n$ = 417), representing roughly one-quarter of cases. There was a trend for a higher percentage of cases involving women to receive scores 1 through 5 compared to cases involving men. More cases involving men received a score of 6. Cases involving men had significantly higher NCA subscale scores with an average of 3.92 ($SD$ = 1.55) compared to cases involving women with an average NCA subscale score of 3.61 ($SD$ = 1.46), although these average scores represented roughly the same level of risk, $t(4{,}536) = 6.44$, $p < .001$, $d = 1.52$.

**Figure 6. Distribution of NCA Subscale Scores by Gender**




## Pretrial Success Rates Across FTA Subscale Scores

Figure 7 shows that across cases included in our analytic sample, most people made it to their court dates. Values in y-axis and above the columns represent the percentage of cases that received each FTA subscale score and the commensurate rates of court appearance. Generally, we would expect court appearance rates to be higher at lower FTA subscale scores and lower at higher FTA subscale scores. Court appearance was higher for FTA subscale scores 1 and 2, with more than two-thirds of cases not having any missed court dates (67.1%, *n* = 342, and 69.1%, *n* = 467, respectively). Cases that received an FTA subscale score of 5 had the lowest appearance rates, but still, in well over half of these cases, people appeared at all their court dates (57.3%, *n* = 499). This pattern of results suggests that the rate of court appearance generally increased appropriately as the level of risk decreased, with some exceptions.

## Figure 7. Distribution of Court Appearance Rates Across FTA Subscale Scores

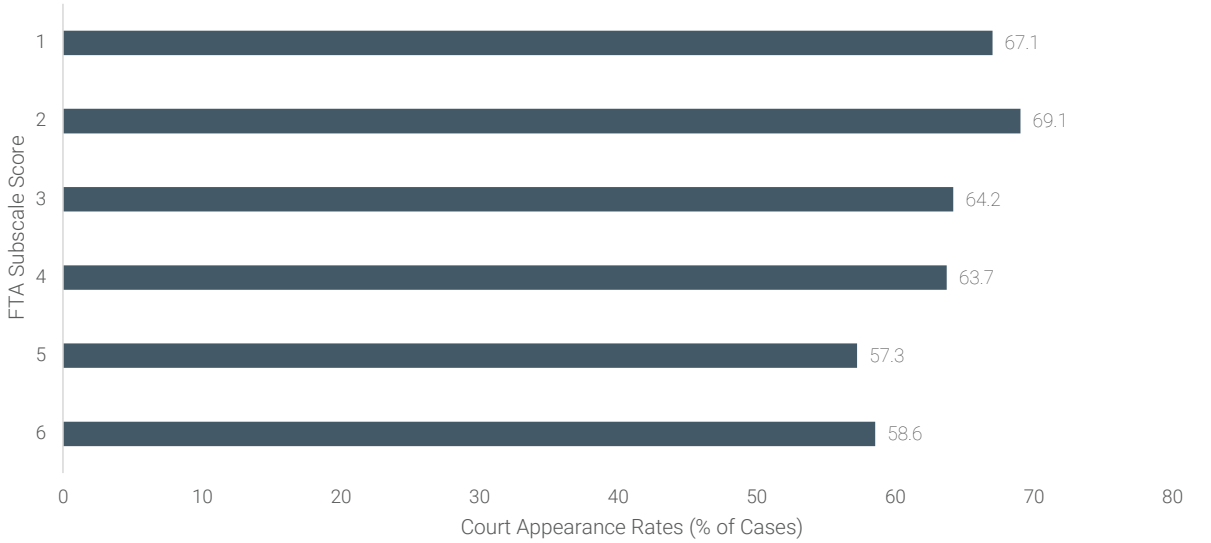



Figure 8 displays different patterns of court appearance rates for Native American compared to white people. For cases involving white people, court appearance rates decrease as FTA subscale scores increase, which is the pattern that we would expect. However, for cases involving Native American people, we see the opposite trend: court appearance rates increase as FTA subscale scores increase. In other words, as the estimated risk of failure to appear increased, the observed rate of failure to appear decreased.

## Figure 8. Court Appearance Rates Across FTA Subscale Scores for Native American and White People

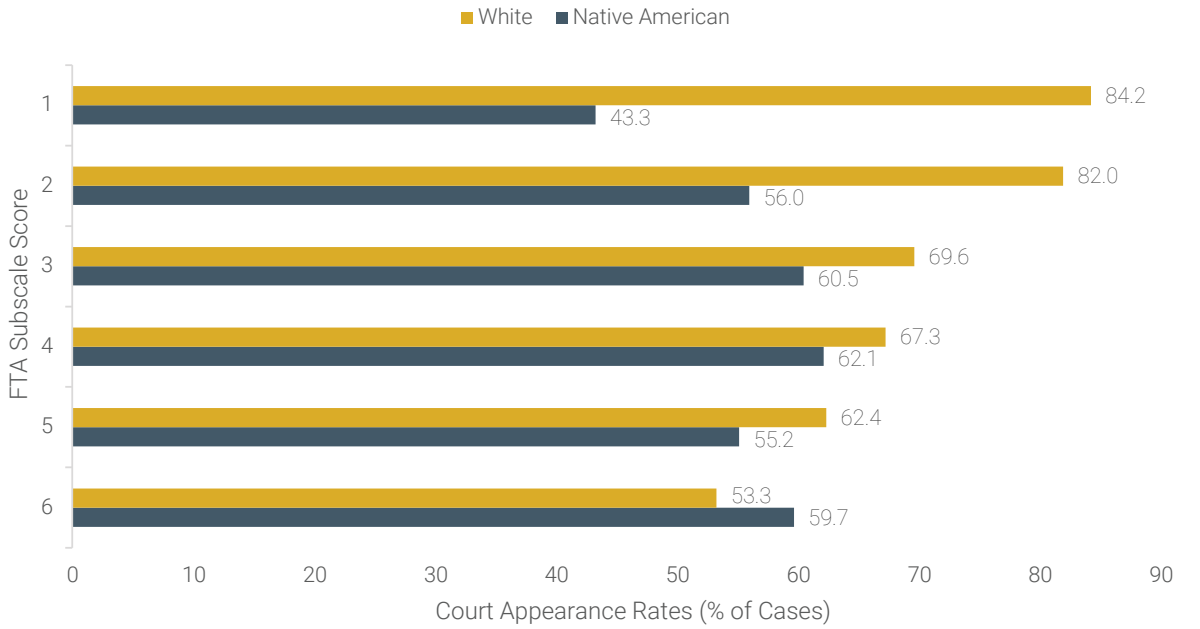
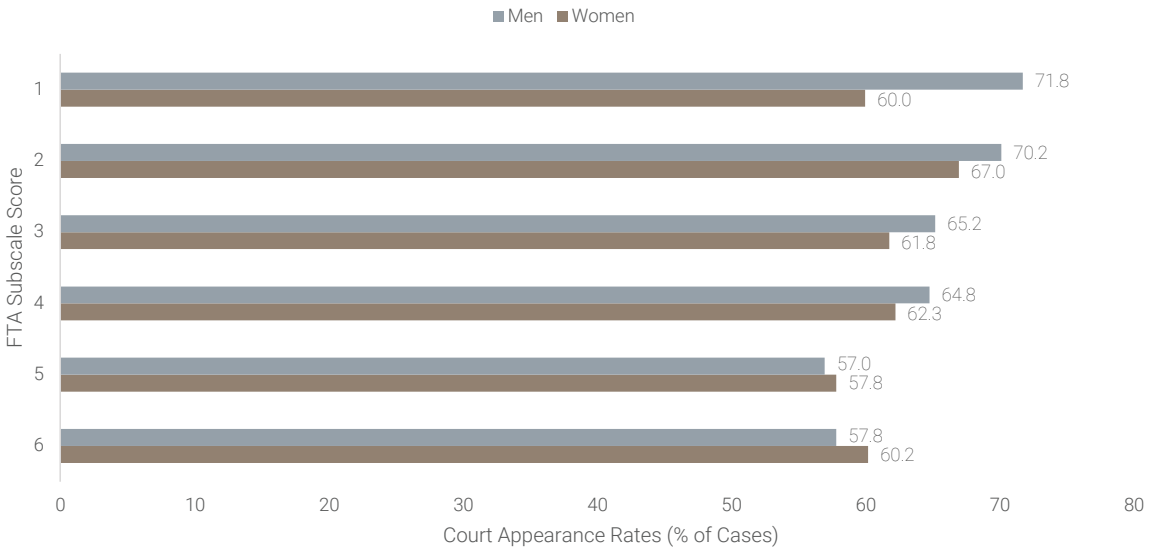


Figure 9 displays comparable patterns of court appearance rates for men and women: generally, we see higher court appearance rates at lower FTA subscale scores and lower court appearance rates at higher FTA subscale scores, which is in keeping with expectations. However, there are some notable exceptions; for example, the rate of court appearance increases (rather than decreases) slightly from FTA subscale scores of 5 to 6 for cases involving men (57.0%, *n* = 310 vs. 57.8%, *n* = 339) and more substantially from FTA subscale scores of 1 to 2 for cases involving women (60.0%, *n* = 111 vs. 67.0%, *n* = 140).

**Figure 9. Court Appearance Rates Across of FTA Subscale Scores for Men and Women**




## Pretrial Success Rates Across NCA Subscale Scores

Figure 10 shows arrest-free rates across all NCA subscale scores during the pretrial period. Values in y-axis and above the columns represent the percentage of cases that received each NCA subscale score and commensurate rates of remaining arrest-free during the pretrial period. Generally, we would expect arrest-free rates to be highest at score 1 and to decrease as scores increase, which is exactly what we see in Figure 10.

## Figure 10. Arrest-Free Rates Across NCA Subscale Scores

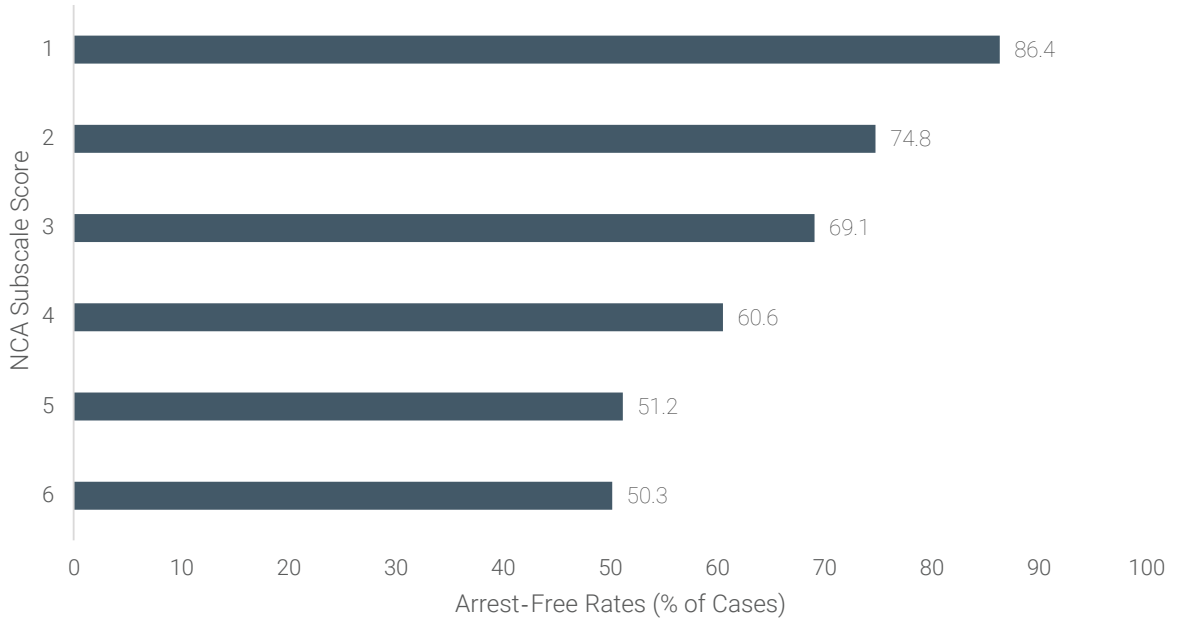



Figure 11 shows different trends in arrest-free rates for Native American and white people. As seen in the overall sample, we found the expected and desired pattern. Specifically, arrest-free rates are the highest at the NCA subscale score of 1 for cases involving both Native American (82.6%, *n* = 95) and white people (87.7%, *n* = 185). Arrest-free rates are the lowest at an NCA subscale score of 6 for case involving Native American people (48.4%, *n* = 296) and 5 for cases involving white people (49.8%, *n* = 104). Arrest-free rates for cases involving white people were slightly higher than those for Native American people at all NCA subscale scores except score 5.

## Figure 11. Arrest-Free Rates Across NCA Subscale Scores for Native American and White People

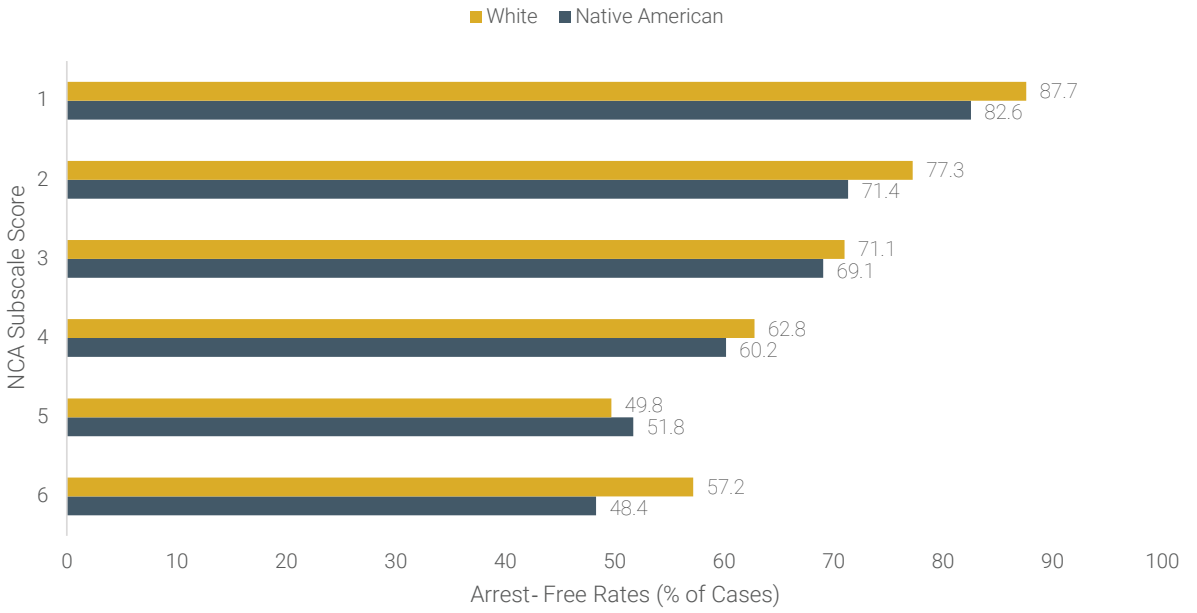



Figure 12 displays similar rates for remaining arrest-free for men and women across NCA subscale scores. Arrest-free rates generally decreased as NCA subscale scores increased for cases involving both men and women with one exception: for cases involving women, arrest-free rates increased from NCA subscale scores of 5 to 6.

## Figure 12. Arrest-Free Rates Across NCA Subscale Scores for Men and Women

## Step 2: Predictive Validity Metrics

Table 7 below presents a summary overview of the findings regarding predictive validity. We summarize the findings using the benchmarks presented in Table 6. We provide a more detailed discussion of the predictive validity of the FTA and NCA subscale scores, overall and by race and gender, in the sections that follow.

### Table 7. Summary of Predictive Validity Results

| ***Failure to Appear*** | ***Acceptability of Predictive Validity*** |
|---|---|
| **Overall** | Poor |
| **Gender** | |
| Men | Fair |
| Women | Poor |
| **Race** | |
| Native American People | Poor |
| White People | Good |
| ***New Criminal Arrest*** | ***Acceptability of Predictive Validity*** |
| **Overall** | Fair |
| **Gender** | |
| Men | Good |
| Women | Fair |
| **Race** | |
| Native American People | Fair |
| White People | Good |

## Failure to Appear

Table 1 in the Statistical Appendix presents the full predictive validity results for the FTA subscale scores. Results show that the FTA subscale scores had poor validity in predicting failure to appear in the overall analytic sample. This means that the FTA subscale score did a poor job of distinguishing between cases with people at a greater and lesser likelihood of failing to appear. Further, results showed that predictive validity differed across race and gender. FTA scores predicted failure to appear with good validity in cases for white people but poor validity in cases for Native American

people; this difference in predictive validity was statistically significant, $Z = 8.07$, $p < .001$. FTA scores predicted failure to appear with fair validity for cases involving men. However, FTA subscale scores did not predict failure to appear for cases involving women at statistically significant levels.

### New Criminal Arrest

Table 1 in the Statistical Appendix presents the full predictive validity results for the NCA subscale scores. Results show that the NCA subscale scores had fair validity in predicting new criminal arrest in the overall analytic sample. In other words, the NCA subscale scores did a fair job of distinguishing between cases with people at a greater and lesser likelihood of being rearrested pretrial. Results also showed that predictive validity differed across race but not across gender. NCA scores predicted rearrest with good validity in cases for white people and fair validity in cases for Native American people; this difference in predictive validity was statistically significant, $Z = 2.30$, $p = .022$. NCA scores predicted failure to appear with good validity for men and fair validity for women; this difference in predictive validity was not statistically significant, $Z = 0.78$, $p = .433$.

## Step 3: Differential Prediction by Race and Gender

### Summary of Differential Prediction Findings

Table 8 below presents a summary overview of the findings related to differential prediction. We provide a more detailed discussion of the differential prediction of the FTA and NCA subscale scores by race and gender in the sections that follow.

**Table 8. Summary of Differential Prediction Results**

| ***Failure to Appear*** | ***Evidence of Differential Prediction*** |
|---|---|
| **Gender** | Yes |
| **Race** | Yes |
| ***New Criminal Arrest*** | ***Evidence of Differential Prediction*** |
| **Gender** | No |
| **Race** | No |

## Failure to Appear

To examine differential prediction for the FTA subscale score, we used multivariable logistic regression models to examine the interactions between race, gender, and the FTA subscale score in predicting failure to appear.

First, we report the results of the model testing for differential prediction by race. Analysis showed that race was statistically associated with the likelihood of failure to appear: cases involving Native American people were almost twice as likely as those involving white people to have a failure to appear (OR = 1.91, $p < .001$). Further, race influenced how well the average FTA subscale score predicted the average likelihood of failing to appear (OR = 0.69, $p < .001$). Specifically, for cases involving white people, rates of failure to appear increased as FTA subscale scores increased while for cases involving Native American people, rates of failure to appear decreased as FTA subscale scores increased. These findings provide evidence of differential prediction, and therefore predictive bias, of FTA subscale scores by race. Table 2 in the Statistical Appendix presents the statistics for the logistic regression model testing for differential prediction of FTA subscale scores by race.

Next, we report the results of the model testing for differential prediction by gender. Analysis showed that gender was not statistically associated with the likelihood of failure to appear (OR = 1.11, $p = .116$). Neither men nor women were more or less likely to fail to appear relative to each other. However, gender did influence how well the average FTA subscale score predicted the average likelihood of failing to appear (OR = 0.904, $p = .012$). Specifically, while rates of failure to appear increased across FTA subscale scores for cases involving both men and women, failure to appear rates are slightly higher for cases involving women at lower FTA subscale scores but slightly higher for cases involving men at higher subscale scores. These findings provide evidence of differential prediction, and therefore predictive bias, of FTA subscale scores by gender. Table 3 in the Statistical Appendix presents the analysis for the logistic regression model testing for differential prediction of FTA subscale scores by gender.

## New Criminal Arrest

To examine differential prediction for the NCA subscale score, we again used multivariable logistic regression models to examine the interactions between race, gender, and the NCA subscale score in predicting new criminal arrest.

First, we report the results of the model testing for differential prediction by race. Analysis showed that race was statistically associated with the likelihood of new

criminal arrests during the pretrial period: cases involving Native American people were about one and a half times more likely than those involving white people to have a failure to appear (OR = 1.51, $p < .001$). Race did not, however, influence how well the average NCA subscale score predicted the average likelihood of new criminal arrest (OR = 0.94, $p = .190$). These findings do not provide evidence of differential prediction--or predictive bias--of NCA subscale scores by race. Table 2 in the Statistical Appendix presents the statistics for the logistic regression model testing for differential prediction of NCA subscale scores by race.

Next, we report the results of the model examining differential prediction by gender. Analysis showed that gender was not statistically associated with the likelihood of new criminal arrest (OR = 0.94, $p = .368$). Further, gender did not influence how well the average NCA subscale score predicted the average likelihood of new criminal arrest (OR = 0.98, $p = .701$). These findings do not provide evidence of differential prediction--or predictive bias--of NCA subscale scores by gender. Table 3 in the Statistical Appendix presents the statistics for the logistic regression model testing for differential prediction of NCA subscale scores by gender.

# Conclusions

## Summary of Findings

The findings regarding the validity of PSA results in predicting pretrial outcomes in Pennington County, SD are mixed. Below, we summarize the report findings and then describe caveats or issues that should be considered in interpreting the results.

When pretrial outcomes were examined in the full sample of people released on their own recognizance, rates of pretrial success generally decreased as subscale scores increased for both court appearances and new criminal arrests. However, FTA subscale scores predicted the likelihood of failure to appear with poor validity overall, a finding apparently driven by the trends seen for Native American people and women. For cases involving Native American people, their rates of court appearance increased as their FTA subscale score increased. For cases involving women, their rates of court appearance were similar across all FTA subscale scores, rather than demonstrating a downward trend as scores increased. Indeed, the FTA subscale score predicted failure to appear significantly better for cases involving white people than Native American people and for cases involving men than women. Race and gender also both significantly influenced how well the average FTA subscale score predicted the average likelihood of failing to appear. Together, results of this validation highlight problems with the validity of the FTA subscale scores overall and provide evidence of bias for cases involving Native American people and women, specifically.

NCA subscale scores predicted likelihood of rearrest during the pretrial period with fair validity (rather than good or excellent) validity. There also were some trends that were inconsistent with expectations. Specifically, for Native American people, their arrest-free rates were highest at subscale scores 3 and 4 (rather than for subscales scores 1 and 2). Though NCA subscale scores predicted new arrests significantly better for white people than for Native American people, race did not significantly influence how well the average NCA subscale score predicted the average likelihood of receiving a new criminal arrest. There were no notable findings in relation to gender. Overall, we found limited evidence of differences in predictive validity and no evidence of differential prediction for the NCA subscale scores.

These findings suggest that validity of the PSA subscale scores in predicting pretrial outcomes in Pennington County ranged from poor to good overall and across race and gender. Findings raise concern regarding the predictive validity of the FTA subscale scores among Native American people and women but not regarding the predictive validity of the NCA subscale scores.

## Limitations

There are some practical limitations to the evaluation data that may affect conclusions based on the findings presented in this report. First, the dataset shared with us did not indicate whether a person was released from jail. Validation studies can only include samples of people who are in the community and thus at risk of experiencing the outcomes of interest during the pretrial period. Thus, we could only examine people released from jail on personal recognizance bonds as this was the only type of release for which we could be certain people were released to the community. We could not examine people who were assigned bail because there was no way for us to know that a person could pay the assigned bail required for their release. As a result, our validation is based on a limited subsample of the total population of people who are assessed using the PSA and released from the jail pretrial in Pennington County.

Second, the dataset did not include the date a person was released from jail or the date a person's case was disposed. Without these dates we were unable to control for the time a person was at risk in our analyses. Yet, the time at risk can influence predictive validity; to demonstrate, a person who spends a longer time in the community pretrial has more time to experience an outcome than who spends less time in the community. Time at risk can be influenced by many factors that can affect the relationship between the PSA scores and the pretrial outcomes.

Third, there may be confounding variables (i.e., external, unrelated influences) that affected our measurement of the outcome variables (failure to appear and new criminal arrest) and, thus, the performance of the PSA subscale scores. For instance, the impact of COVID on criminal legal practices may have affected the frequency of pretrial outcomes observed for some people (e.g., courts may have stopped logging non-appearance rates, officers may not have arrested people for behaviors for which people would have been arrested prior to COVID). As a result, findings should be interpreted in light of additional potential influences on pretrial outcomes.

## Recommendations

1. **Limit the use of the FTA subscale scores to inform decisions about service delivery.** Our findings indicate that the FTA subscale scores have poor predictive validity and exhibit racial and gender bias, making them unsuitable for use in pretrial decision-making for the sample we analyzed. We recommend that FTA subscale scores not be used to determine pretrial release conditions. If considered, FTA subscale scores should only inform voluntary services aimed

at supporting court appearances. These services should not be mandated as a condition of release but rather offered to enhance court compliance without contributing to biased pretrial decisions.

2. **Consider adopting an instrument that assesses dynamic factors.** The PSA is limited to static (i.e., unchanging), largely historical items that emphasize prior criminal legal system contact.[46] Yet research in other risk assessment contexts finds that more proximal and dynamic items (i.e., recent, changing, or modifiable) add incremental predictive validity and demonstrate less predictive bias than static factors.[47,48] Further, a recent synthesis of research on the performance of risk assessment instruments in Indigenous populations specifically found that dynamic risk assessment instruments performed better than static risk assessment instruments in this population.[49] We recommend completing an assessment of dynamic factors along with the PSA to provide additional information to inform pretrial release decisions and conditions, as appropriate. Instruments to consider could include the Risk and Needs Triage (RANT)[50] or the Personal Recognizance Interview & Needs Screen (PRINS).[51] A mental health or substance use screening instrument could also provide information on dynamic factors to inform decisions related to supportive services. Examples of validated and free screening tools include the Brief Jail Mental Health Screen (BJMHS),[52] the Texas Christian University Drug Screen 5 (TCUDS-V)[53] or Alcohol Use Disorders Identification Test (AUDIT),[54] and the PTSD Checklist for DSM-5 (PCL-5).[55] Web-based screening tools could be hosted on the shared data dashboard, and policies should allow the screening results to be shared with appropriate partners in a timely fashion for case planning.
3. **Clarify override procedures and monitor their use and impact on decision-making.** At initial appearance hearings, judges in Pennington County can override PSA subscale scores. Overrides occur when judges make pretrial release decisions that do not align with the supervision level indicated by the PSA decision-making matrix. Yet, research suggests that judicial overrides can decrease predictive validity and contribute to higher rates of pretrial misconduct.[56,57] We recommend evaluating how often judges use overrides, for whom, and why to help to explain the differences in the predictive validity of the PSA across race and gender. This information could inform strategies to support successful—and equitable—implementation of the PSA in practice.

4. **Implement continuous quality improvement (CQI) measures.** Consistent with national guidelines,[58] we recommend the development and implementation of CQI measures, including a twice yearly case review of fidelity, consistency, and consistency of the PSA scoring process, and routine booster training to refresh PSA users on administration procedures and to address issues as they arise. CQI measures could include the development of a case review checklist that includes observable indicators of fidelity and consistency of PSA scores, as well as steps in the assessment process. The checklist could include looking for missing ratings, ensuring ratings are updated as needed (e.g., if a person's charge information changes), inconsistencies between item scores and risk levels, or inconsistencies in how information is collected and scored for different people. The Risk Assessment Quality Improvement (RAQI) protocol could provide some initial ideas for what to include in such a case review checklist and for how to establish a quality improvement process, more generally. Additionally, Advancing Pretrial Policy and Research has an implementation guide to quality assurance that is available for free download. The CQI process should be ongoing to ensure the PSA is being completed accurately.
5. **Conduct a validation with a more representative sample and additional data.** The current validation focused on people released on personal recognizance bonds and whose cases were closed at the time of data extraction. There also was information missing that would have improved our statistical analyses (e.g., information on time at risk). We recommend another validation be conducted on a broader sample and with additional data to provide more comprehensive and generalizable findings regarding the performance of the PSA subscale scores in Pennington County, SD. Additional data should include:
    - **New violent criminal arrests during the pretrial period.** Currently, staff record whether a person receives any new criminal arrest during the pretrial period. Staff must also record whether a person receives a new violent arrest specifically. A record of new violent arrests is necessary to evaluate the predictive validity of the NVCA subscale.
    - **Date of release from jail.** To be included in a validation sample, a person must be released from jail and in the community for at least some portion of their pretrial period. Staff must record the date a person is released from jail so evaluators can confirm a person has been released (and was thus at risk for pretrial outcomes) and calculate a person's total time in the community.

- **Date of case disposition.** The PSA is only intended to predict outcomes during the pretrial period (i.e., between arrest and case disposition). Staff must record the date a person's case is disposed so researchers can calculate a person's total time in the community during the pretrial period.
- **All pretrial outcomes experienced.** Currently, staff can only record one instance of either failure to appear in court or new criminal arrest. The first instance of failure represents the end point in the current data. Yet, presumably, a person could both fail to appear in court *and* later receive a new criminal arrest. Creating a system that would allow staff to record all possible outcomes would increase the sample size that can be used to validate each PSA subscale score and would contribute to more accurate and externally valid results.

# References


1 VanNostrand, M., & Lowenkamp, C. T. (2013). *Assessing pretrial risk without a defendant interview.*

2 Desmarais, S. L., & Lowder, E. M. (2019). *Pretrial risk assessment tools: A primer for judges, prosecutors, and defense attorneys.* Safety and Justice Challenge.

3 Ibid.

4 Ibid.

5 *About the public safety assessment.* (n.d.). Advancing Pretrial Policy & Research (APPR). https://advancingpretrial.org/psa/about/

6 VanNostrand, M., & Lowenkamp, C. T. (2013).

7 Hess, J., & Turner, S. (2021). *Validation of the PSA in Los Angeles County.*

8 Skog, A., & Lacoe, J. (2021). *Validation of the PSA in San Francisco.*

9 Brittain, B. J., Georges, L., & Martin, J. (2021). Examining the predictive validity of the Public Safety Assessment. *Criminal Justice and Behavior, 48*(10), 1431–1449.

10 DeMichele M., Tueller, S., Inkpen, C., Dawes, D., & Lattimore, P. K. (2023). *Validation and predictive bias testing of the Public Safety Assessment for Fulton County, Georgia.*

11 Greiner, D. J., Stubenberg, M., & Halen, R. (2020). *Validation of the PSA in McLean County, IL.*

12 Greiner, D. J., Stubenberg, M., & Halen, R. (2021). *Validation of the PSA in Kane County, IL.*

13 VanNostrand, M., & Lowenkamp, C.T. (2013).

14 DeMichele, M., Baumgartner, P., Wenger, M., Barrick, K., & Comfort, M. (2020). Public safety assessment: Predictive utility and differential prediction by race in Kentucky. *Criminology & Public Policy, 19*(2), 409–431

15 Lowenkamp C., DeMichele, M., & Klein Warren, L. (2020). *Replication and extension of the Lucas County PSA project.*

16 Greiner, D.J., Stubenberg, M., & Halen, R. (2020). *Validation of the PSA in Harris County, TX.*

17 DeMichele, M., Tueller, S., Janda, K., Dawes, D., & Lattimore, P. K. (2023). *Validation and predictive bias testing of the Public Safety Assessment for Thurston County, Washington.*

18 DeMichele, M., Tueller, S., Burtch, E., Dawes, D., & Lattimore, P. K. (2023). *Validation and predictive bias testing of the Public Safety Assessment for Pierce County, Washington.*

19 Skog, A., & Lacoe, J. (2021).

20 DeMichele, M., Baumgartner, P., Wenger, M., Barrick, K., & Comfort, M. (2020).

21 DeMichele, M., Tueller, S., Janda, K., Dawes, D., & Lattimore, P. K. (2023).

22 DeMichele, M., Tueller, S., Burtch, E., Dawes, D., & Lattimore, P. K. (2023).

23 DeMichele, M., Tueller, S., Inkpen, C., Dawes, D., & Lattimore, P. K. (2023).

24 Greiner, D. J., Stubenberg, M., & Halen, R. (2020).

25 Greiner, D. J., Stubenberg, M., & Halen, R. (2021).

26 DeMichele, M., Baumgartner, P., Wenger, M., Barrick, K., & Comfort, M. (2020).

27 Lowenkamp, C., DeMichele, M., & Klein Warren, L. (2020).

28 Greiner, D. J., Stubenberg, M., & Halen, R. (2020).

29 DeMichele, M., Tueller, S., Janda, K., Dawes, D., & Lattimore, P. K. (2023).

30 DeMichele, M., Tueller, S., Burtch, E., Dawes, D., & Lattimore, P. K. (2023).

31 Skog, A., & Lacoe, J. (2021).

32 DeMichele, M., Baumgartner, P., Wenger, M., Barrick, K., & Comfort, M. (2020).

33 DeMichele, M., Tueller, S., Janda, K., Dawes, D., & Lattimore, P. K. (2023).

34 DeMichele, M., Tueller, S., Burtch, E., Dawes, D., & Lattimore, P. K. (2023).

35 DeMichele, M., Tueller, S., Inkpen, C., Dawes, D., & Lattimore, P. K. (2023).

36 Greiner, D.J., Stubenberg, M., & Halen, R. (2020).

37 Greiner, D.J., Stubenberg, M., & Halen, R. (2021).

38 DeMichele, M., Baumgartner, P., Wenger, M., Barrick, K., & Comfort, M. (2020).

39 Lowenkamp, C., DeMichele, M., & Klein Warren, L. (2020).

40 Greiner, D. J., Stubenberg, M., & Halen, R. (2020).

41 DeMichele, M., Tueller, S., Janda, K., Dawes, D., & Lattimore, P. K. (2023).

42 DeMichele, M., Tueller, S., Burtch, E., Dawes, D., & Lattimore, P. K. (2023).

43 Brittain, B. J., Georges, L., & Martin, J. (2021).

44 DeMichele, M., Baumgartner, P., Wenger, M., Barrick, K., & Comfort, M. (2020).

45 Desmarais, S. L., Johnson, K. L., & Singh, J. P. (2016). Performance of recidivism risk assessment instruments in US correctional settings. *Psychological Services, 13*(3), 206.

46 VanNostrand, M., & Lowenkamp, C. T. (2013).

47 Desmarais, S. L., Nicholls, T. L., Wilson, C. M., & Brink, J. (2012). Using dynamic risk and protective factors to predict inpatient aggression: Reliability and validity of START assessments. *Psychological Assessment, 24,* 685–700.

48 Johnson, K. L., Desmarais, S. L., Grimm, K. J., Tueller, S., Swartz, M. S., & Van Dorn, R. A. (2016). Proximal risk factors for short-term community violence among adults with mental illnesses. *Psychiatric Services, 67,* 771–778.

49 Olver, M. E., Stockdale, K. C., Helmus, L. M., Woods, P., Termeer, J., & Prince, J. (2024). Too risky to use, or too risky not to? Lessons learned from over 30 years of research on forensic risk assessment with Indigenous persons. *Psychological Bulletin.*

50 Marlowe, D. B., Festinger, D. S., Dugosh, K. L., Caron, A., Podkopacz, M. R., & Clements, N. T. (2011). Targeting dispositions for drug-involved offenders: A field trial of the Risk and Needs Triage (Rant)™. *Journal of Criminal Justice, 39*(3), 253–260.

51 Hamilton, Z., Ursino, J., & Kigerl, A. (2024). Advancing pretrial assessments: Development of the personal recognizance interview & needs screen(Prins). *Journal of Criminal Justice, 92.*

52 Steadman, H. J., Scott, J. E., Osher, F., Agnese, T. K., & Robbins, P. C. (2005). Validation of the brief jail mental health screen. *Psychiatric Services, 56*(7), 816–822.

53 *TCU drug screen V (TCUDS V).* (2014). Texas Christian University..

54 Saunders, J. B., Aasland, O. G., Babor, T. F., De la Fuente, J. R., & Grant, M. (1993). Development of the alcohol use disorders identification test (Audit): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction, 88*(6), 791–804.

55 Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The posttraumatic stress disorder checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress, 28*(6), 489–498.

56 Bonta, J., & Andrews, D. A. (2007). Risk-need-responsivity model for offender assessment and rehabilitation. *Rehabilitation, 6*(1), 1–22.

57 VanNostrand, M., & Keebler, G. (2009). Pretrial risk assessment in the federal court. *Federal Probation, 73*(2), 3–29.

58 Desmarais, S. L., D'Amora, D. A., & Tavárez, L. P. (2022). *Advancing fairness and transparency: National guidelines for post-conviction risks and needs assessment.* Council of State Governments Justice Center.

# Statistical Appendices

## Statistical Appendix 1: AUCs for FTA and NCA Subscales Predicting their Respective Outcomes

| Sample | Outcomes | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Failure to Appear | | | | | New Criminal Arrest | | | | |
| | | | | 95% CI | | | | | 95% CI | |
| | AUC | SE | p | LL | UL | AUC | SE | p | LL | UL |
| **Full Sample** | .55 | .01 | <.001 | 0.53 | 0.56 | .63 | .01 | <.001 | 0.62 | 0.65 |
| **Native American** | .48 | .01 | 0.011 | 0.46 | 0.50 | .61 | .01 | <.001 | 0.59 | 0.63 |
| **White** | .64 | .02 | <.001 | 0.61 | 0.67 | .65 | .02 | <.001 | 0.62 | 0.68 |
| **Men** | .56 | .01 | <.001 | 0.54 | 0.58 | .64 | .01 | <.001 | 0.62 | 0.66 |
| **Women** | .52 | .02 | 0.243 | 0.49 | 0.55 | .62 | .02 | <.001 | 0.59 | 0.65 |

*Notes.* Full sample *n* = 4570, Native American *n* = 2876, white *n* = 1454, men *n* = 3015, women *n* = 1523

AUC = area under the curve

SE = standard error

LL = lower limit

UL = upper limit

# Statistical Appendix 2: Results of Logistic Regression Models Predicting Failure to Appear and New Criminal Arrest by Race

| | *Model Statistics* | | | | | | |
|---|---|---|---|---|---|---|---|
| ***Predictors*** | | | | | | ***95% CI*** | |
| ***Failure to Appear*** | ***Estimate*** | ***SE*** | ***Wald*** | ***p*** | ***OR*** | ***LL*** | ***UL*** |
| ***Step 1*** | | | | | | | |
| Constant | -0.96 | .06 | 267.70 | <.001 | 0.38 | -- | -- |
| Race (White) | 0.65 | .07 | 85.87 | <.001 | 1.91 | 1.66 | 2.19 |
| ***Step 2*** | | | | | | | |
| Constant | -1.13 | .09 | 167.30 | <.001 | 0.32 | -- | -- |
| Race (White) | 0.59 | .07 | 66.59 | <.001 | 1.81 | 1.57 | 2.08 |
| FTA subscale score | 0.05 | .02 | 7.03 | .008 | 1.06 | 1.01 | 1.10 |
| ***Step 3*** | | | | | | | |
| Constant | -1.97 | .14 | 189.32 | <.001 | 0.14 | -- | -- |
| Race (White) | 1.89 | .18 | 111.77 | <.001 | 6.62 | 4.66 | 9.39 |
| FTA subscale score | 0.31 | .04 | 66.53 | <.001 | 1.36 | 1.27 | 1.47 |
| Race by FTA subscale score | -0.37 | .05 | 65.75 | <.001 | 0.69 | 0.64 | 0.76 |

| ***Predictors*** | | | | | | ***95% CI*** | |
|---|---|---|---|---|---|---|---|
| ***New Criminal Arrest*** | ***Estimate*** | ***SE*** | ***Wald*** | ***p*** | ***OR*** | ***LL*** | ***UL*** |
| ***Step 1*** | | | | | | | |
| Constant | -0.80 | .06 | 200.11 | <.001 | 0.45 | -- | -- |
| Race (White) | 0.41 | .07 | 36.21 | <.001 | 1.51 | 1.32 | 1.72 |

| *Predictors* | | | | | | *95% CI* | |
|---|---|---|---|---|---|---|---|
| ***New Criminal Arrest*** | ***Estimate*** | ***SE*** | ***Wald*** | ***p*** | ***OR*** | ***LL*** | ***UL*** |
| ***Step 2*** | | | | | | | |
| Constant | -1.84 | .10 | 350.00 | <.001 | 0.16 | -- | -- |
| Race (White) | 0.16 | .07 | 5.21 | .022 | 1.18 | 1.02 | 1.36 |
| NCA subscale score | 0.31 | .02 | 184.33 | <.001 | 1.36 | 1.30 | 1.42 |
| ***Step 3*** | | | | | | | |
| Constant | -1.98 | .15 | 181.64 | <.001 | 0.14 | -- | -- |
| Race (White) | 0.40 | .19 | 4.26 | .039 | 1.49 | 1.02 | 2.18 |
| NCA subscale score | 0.35 | .04 | 82.97 | <.001 | 1.42 | 1.31 | 1.53 |
| Race by NCA subscale score | -0.06 | .05 | 1.72 | .190 | 0.94 | 0.86 | 1.03 |

*Notes. N* = 4570. The reference category is white.

SE = standard error

OR = odds ratio

LL = lower limit

UL = upper limit

# Statistical Appendix 3: Results of Logistic Regression Models Predicting Failure to Appear and New Criminal Arrest by Gender

| | *Model Statistics* | | | | | | |
|---|---|---|---|---|---|---|---|
| ***Predictors*** | | | | | | ***95% CI*** | |
| ***Failure to Appear*** | ***Estimate*** | ***SE*** | ***Wald*** | ***p*** | ***OR*** | ***LL*** | ***UL*** |
| ***Step 1*** | | | | | | | |
| Constant | -0.56 | .04 | 218.87 | <.001 | 0.57 | -- | -- |
| Gender (men) | 0.10 | .07 | 2.47 | .116 | 1.11 | 0.98 | 1.26 |
| ***Step 2*** | | | | | | | |
| Constant | -0.95 | .08 | 131.99 | <.001 | 0.39 | -- | -- |
| Gender (men) | 0.10 | .07 | 2.40 | .121 | 1.11 | 0.97 | 1.26 |
| FTA subscale score | 0.10 | .02 | 28.55 | <.001 | 1.11 | 1.07 | 1.15 |
| ***Step 3*** | | | | | | | |
| Constant | -1.08 | .10 | 119.17 | <.001 | 0.34 | -- | -- |
| Gender (men) | 0.49 | .17 | 8.55 | .003 | 1.63 | 1.17 | 2.25 |
| FTA subscale score | 0.14 | .02 | 33.38 | <.001 | 1.15 | 1.10 | 1.20 |
| Gender by FTA subscale score | -0.10 | .04 | 6.32 | .012 | 0.904 | 0.84 | 0.98 |

| ***Predictors*** | | | | | | ***95% CI*** | |
|---|---|---|---|---|---|---|---|
| ***New Criminal Arrest*** | ***Estimate*** | ***SE*** | ***Wald*** | ***p*** | ***OR*** | ***LL*** | ***UL*** |
| ***Step 1*** | | | | | | | |
| Constant | -0.50 | .04 | 179.29 | <.001 | 0.61 | -- | -- |
| Gender (men) | -0.06 | .07 | .81 | .368 | 0.94 | 0.83 | 1.07 |
| ***Step 2*** | | | | | | | |
| Constant | -1.80 | .10 | 344.36 | <.001 | 0.165 | 1.33 | 1.44 |
| Gender (men) | 0.04 | .07 | .40 | .528 | 1.04 | 0.92 | 1.19 |
| NCA subscale score | 0.32 | .02 | 223.31 | <.001 | 1.38 | -- | -- |

| ***Predictors*** | | | | | | ***95% CI*** | |
|---|---|---|---|---|---|---|---|
| ***New Criminal Arrest*** | ***Estimate*** | ***SE*** | ***Wald*** | ***p*** | ***OR*** | ***LL*** | ***UL*** |
| ***Step 3*** | | | | | | | |
| Constant | -1.83 | .11 | 255.21 | <.001 | 0.16 | -- | -- |
| Gender (men) | 0.11 | .19 | .34 | .563 | 1.12 | 0.77 | 1.64 |
| NCA subscale score | 0.330 | .03 | 159.00 | <.001 | 1.39 | 1.32 | 1.46 |
| Gender by NCA subscale score | -0.02 | .047 | .147 | .701 | 0.98 | 0.90 | 1.08 |

*Notes. N* = 4570. The reference category is men.

SE = standard error

OR = odds ratio

LL = lower limit

UL = upper limit