

Validation and Predictive Bias Testing of the Public Safety Assessment for Pierce County, Washington

Matthew DeMichele, RTI International
Stephen Tueller, RTI International
Emily Burtch, RTI International
Debbie Dawes, RTI International
Pamela K. Lattimore, RTI International

Table of Contents

Executive Summary	4
Introduction	5
What is a Pretrial Assessment?	5
Why is it important to Validate Pretrial Assessments?	5
Characteristics of People Admitted into the Pierce County Jail	6
How do PSA-eligible individuals score on the PSA	8
Validation of the Public Safety Assessment	12
Validation Sample: 4,111 Released Individuals	12
Predictive Validity	14
Assessing Predictive Bias: Ensuring Equal Probabilities	15
Calibration: Equal Rates Across Groups	17
PSA Pretrial Outcomes for Race and Sex Subgroups	19
Conclusion	25

Exhibits

Figure 1: Validation Sample with Negative Outcomes by Race	17
Figure 2: Validation Sample with Negative Outcomes by Sex	18
Figure 3: Plots of Predicted Probabilities for Pretrial Outcomes by Race	21
Figure 4: Plots of Predicted Probabilities for Pretrial Outcomes by Sex	23
Table 1: Descriptive Characteristics by Release Status	6
Table 2: Most Serious Charge for Individuals Booked into the Pierce County Jail (2017-2018)	7
Table 3: Risk Factors by Release Status	8
Table 4: Average PSA Scores by Release Status	10
Table 5: Scale Scores by Released and Detained Status	11
Table 6: Scale Score Distribution in Pierce County	12
Table 7: Negative Pretrial Outcomes for Released Individuals by Scale Score	13
Table 8: Predictive Validity Results: Area Under the Curve (AUC) Values	14
Table 9: Base Rates for FTA, NCA, and NVCA by Race	16
Table 10: Base Rates for FTA, NCA, and NVCA by Sex	16
Table 11: AUC Values by Race and Sex	16
Table 12: Logistic Regression Results Testing for Predictive Bias by Race	20
Table 13: Logistic Regression Results Testing for Predictive Bias by Sex	22

Acknowledgments

The authors would like to thank individuals in Pierce County for their willingness to share data, and their time, and expertise to support RTI's research. We especially thank Andrea Kelly, Sarah Schaub, and Andrew Peterson. We thank Arnold Ventures for their support of this research and extend special appreciation to Kristin Bechtel and Virginia Bersch for their insights, feedback, and direction. This project involves a large team effort that has benefitted from support by RTI's Racial Community and Justice Committee members Megan Comfort, Elizabeth Tibaduiza, Jenn Rineer, and Monica Sheppard. Megan Nyce deserves special thanks for her contributions editing and formatting several drafts of this manuscript. Any errors and all points of view are the sole responsibility of the authors.

Executive Summary

In this report, we present findings from a validation and predictive bias test of the Public Safety Assessment (PSA) in Pierce County, Washington. Pierce County is committed to criminal legal system improvements that limit the use of jail incarceration, especially the unnecessary use of pretrial detention. The jurisdiction is a member of Arnold Ventures' Advancing Pretrial Policy and Research (APPR) initiative, through which they have received Training and Technical Assistance (TTA) from a team of experts led by the Center for Effective Public Policy (CEPP) and research support from RTI International. As a part of APPR, Pierce County officials are engaged in efforts across their local criminal legal system to develop improvements to their pretrial system. These improvements are intended to enhance pretrial decision making, reduce pretrial detention, and conduct research to identify and reduce racial disparities. An assessment tool that holds promise for facilitating such advancements is the PSA, which has been associated with reductions in missed court appearances and new crimes, and fewer jail admissions.¹ RTI is contributing to Pierce County's recent efforts to improve local pretrial practices through analyses to determine whether the PSA is a tool that is valid for their jurisdiction.

The PSA provides court actors with information about the likelihood (predicted probability) that individuals with specific characteristics will miss a court appearance (fail to appear, FTA), be arrested for a new crime (new criminal arrest, NCA), and be arrested for a new violent crime (new violent criminal arrest, NVCA) if they are released pretrial. The PSA consists of 9 factors used across three scales to measure likelihood of FTA, NCA, and NVCA. The three PSA scores (FTA, NCA, NVCA) range in value from 1 to 6, with 1 indicating the lowest likelihood and 6 the highest likelihood for each of the outcomes.

The current report focuses on a validation and predictive bias testing of the PSA using data about adults admitted into Pierce County's jail and adjudicated in Superior Court for a new felony charge between January 1, 2017 and December 31, 2018. The results demonstrate that the PSA meets standards of predictive validity for criminal legal system assessments.² Higher scores are related to significantly greater likelihood that someone will miss court or be rearrested for a new crime during their pretrial release.

We did not find any evidence of predictive bias by race or sex for the PSA in Pierce County. The predictions from the PSA for FTA, NCA, and NVCA are not different for White individuals and people of color or males and females in the historical Pierce County data.

The current results show that the PSA is associated with pretrial outcomes, and we did not find evidence that the PSA exacerbates predictive bias related to race and sex. Pretrial assessment research is ongoing and requires using new local data to compare the current historical findings to future outcomes post-PSA implementation to avoid misstating the likelihood of outcomes.³ Proper use of assessments requires ongoing research to ensure that the PSA remains valid for Pierce County.

1 Lowenkamp, C., DeMichele, M., and Klein Warren, L. (2020). Replication and extension of the Lucas County PSA project. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3727443 The authors found the post-PSA period was associated with about 1,590 fewer bookings each year.

2 Desmarais, S., Zottola, S., Clarke, S., and Lowder, E. (2021). Predictive validity of pretrial risk assessments: A systematic review of the literature. *Criminal Justice and Behavior*, 48(4): 398-420.

3 Koepke, Logan, J., & Robinson, D. G. (2018). *Danger Ahead: Risk Assessment and the Future of Bail Reform*, Retrieved from <https://ssrn.com/abstract=3041622> or <https://doi.org/10.2139/ssrn.3041622>

Introduction

This report presents findings from a validation and predictive bias study to assess the validity of the Public Safety Assessment (PSA) for application in pretrial assessments for individuals booked into jail in Pierce County, Washington. The validation study uses historic data to examine how the PSA would have performed in a recent period to determine whether the PSA is valid for implementation in Pierce County.

What is a Pretrial Assessment?

A pretrial assessment is a tool to inform the decisions made by judicial officers in a pretrial setting. Pretrial assessments inform decisions about release and supervision conditions by estimating the statistical likelihood that an individual with specific characteristics will be charged with a new crime (NCA and NVCA) and will miss a court appointment (FTA) if they are released pretrial. The PSA uses nine factors to estimate the probability of three pretrial outcomes: failure to appear in court (FTA), an arrest for a new crime (NCA), and an arrest for a new violent crime (NVCA). Most of the PSA's factors are related to an individual's prior criminal convictions, prior FTAs, whether the current charge is for a crime considered violent, and their age.

Pretrial assessments provide a basis for court officials to incorporate jurisdiction-specific recommendations about release conditions based on the level of risk predicted by the assessment tool. Assessments do not determine whether an individual should be released or detained pretrial. Rather, assessment scores are one of several sources of information for pretrial decision-makers can consider when setting pretrial conditions. Valid, unbiased pretrial assessments can contribute to goals of pretrial justice, which include maximizing public safety, court appearance, and pretrial release.

Why is it Important to Validate Pretrial Assessments?

Validation of the PSA is important because any assessment tool must provide a good and unbiased “fit” to the local context and population to which it is applied. An assessment that predicts pretrial outcomes with a high degree of accuracy in one jurisdiction may not perform well in another. Thus, it is critically important to evaluate how well an assessment measures the probability of pretrial outcomes through a process known as validation. Validations use local historical data to determine whether the assessment is accurate in terms of predicting outcomes for the local population and bias testing assesses the extent to which the assessment does not over- or under-predict for racial groups or for females compared to males.

Pretrial assessments should be revalidated regularly to account for shifts in local legal policy and demographics. In addition, large-scale events – like the COVID-19 pandemic – that impact pretrial processes and outcomes may necessitate revalidation. Revalidation ensures that the assessment reflects local pretrial outcomes, advances the policy goals of local officials, and does not exhibit predictive bias.

Characteristics of People Admitted into the Pierce County Jail

The data for our study include 6,437 jail bookings that occurred in Pierce County, Washington, between January 1, 2017, and December 31, 2018, for new cases assigned to Superior Court. Table 1 provides descriptive characteristics of individuals booked into Pierce County jail by their release or detention status. An individual was defined as “released” if they were released at any point prior to the final disposition of their case. An individual was defined as “detained” if they spent their entire pretrial period in jail. In Pierce County, about one-third (N = 2,326, 36%) of the individuals admitted to jail were detained and about two-thirds were released (N = 4,111, 64%). Pierce County is implementing the PSA for individuals who are charged with felony crimes and whose cases are referred to Superior Court. Thus, the validation for Pierce County focuses on individuals charged with felony crimes and referred to Superior Court.

Table 1: Descriptive Characteristics by Release Status

	Detained (N=2,326)	Released (N=4,111)	Overall (N=6,437)	2019 Census (N=904,980) [§]
American Indian/ Alaska Native	102 (4.4%)	107 (2.6%)	209 (3.2%)	1.8%
Asian/Pacific Islander	140 (6.0%)	297 (7.2%)	437 (6.8%)	8.9%
Black	618 (26.6%)	914 (22.2%)	1,532 (23.8%)	7.7%
Two or More Races	NA*	NA*	NA*	7.4%
Unknown	28 (1.2%)	69 (1.7%)	97 (1.5%)	NA†
White	1,438 (61.8%)	2,724 (66.3%)	4,162 (64.7%)	65.7%

* This category is not defined in the Pierce County jail admission data

† This category is not defined in the census report

§ The 2019 Census estimates are for the entire population (including children) while the PSA eligible

About two-thirds of the validation sample are White individuals⁴ (N = 4,162). Nearly one-quarter of the individuals booked are Black (N = 1,532, 24%), 7% (N = 437) are Asian or Pacific Islander, and 3% (N = 209) are American Indian or Alaska Native. The breakdowns by racial categories do not show large differences in the release and detention status among the various groups. For instance, Black individuals are 24% of the overall booked population and 22% of the released population. However, Black individuals are disproportionately booked relative to their presence in Pierce County as 24% of booked individuals are Black relative to their 8% presence in the general population.

⁴ Our categorization of people's racial and ethnic identities relies on the categories used in the Pierce County criminal legal system data. In this report, we categorize people as “White” if that is only racial/ethnic identifier available for them. People who are identified as belonging to all other racial/ethnic categories are categorized as “people of color.” For more on the complexities of these issues, see [Using Data to Explore Racial and Ethnic Disparities in the Criminal Legal System](#).

Table 2 provides information on the most serious charge for the 6,437 people charged with at least one felony offenses and booked into the Pierce County jail between January 1, 2017, and December 31, 2018. In order of prevalence, the most serious charge associated with a booking was classified as a drug (16%), other (0.1%), property (39%), public order (9%), or violent (37%).⁵

On average, people booked into Pierce County jail spent 47 days in jail. People who were released spent an average of 10 days in jail, substantially shorter than the average of 112 days in jail for those who were detained. Nearly 50% of the people released pretrial, were released within 2 days.

Table 2: Most Serious Charge for Individuals Booked into the Pierce County Jail (2017-2018)

Most serious charge overall category	Total Booked N (percentage)	Released N (percentage)
Drug Offenses	1003 (15.58%)	738 (17.95%)
Other Offenses	2 (0.03%)	1 (0.02%)
Property Offenses	2,502 (38.87%)	1711 (41.62%)
Public Order Offenses	567 (8.81%)	428 (10.41%)
Violent Offenses	2,363 (36.71%)	1233 (29.99%)
Total	6,437 (100%)	4,111 (100%)

⁵ Charge categories are derived from the Bureau of Justice Statistics (BJS) National Corrections Reporting Program broad charge categories using the most serious offense charged for everyone.

How do PSA-eligible individuals score on the PSA

The PSA includes 9 factors that contribute to the three PSA outcome scores (FTA, NCA, and NVCA). RTI obtained data to calculate the factors by linking administrative records from the Pierce County jail and courts and Washington’s statewide criminal history repository. Table 3 shows the proportion of PSA-eligible individuals who have each of the PSA factors at the time of their booking. Each scale includes between 4 and 7 factors, as indicated in the rightmost three columns of Table 3. Scale scores range from 1 to 6 with 1 indicating the lowest and 6 indicating the highest probability of experiencing the outcomes. In this report, when we refer to the outcomes, we are referring to probability of someone experiencing an FTA, NCA, or NVCA.

The PSA was created through investments made by Arnold Ventures using a large database drawn from several jurisdictions to examine the predictive validity of hundreds of risk factors.⁶ The PSA was developed to identify the strongest predictors of failure to appear (FTA), new criminal activity (NCA), and new violent criminal activity (NVCA). The PSA leaves out demographic factors related to race, ethnicity, and sex as well as socioeconomic variables such as residential stability, educational attainment, and employment. These items were excluded to reduce potential for predictive bias for those in lower socioeconomic brackets and people of color. The PSA is completed with court and criminal history data and does not require an interview, and RTI can score the assessment by linking administrative records from the jail, courts, and Washington’s statewide criminal history repository. The PSA is available to the public and jurisdictions can use the PSA for free.

Table 3: Risk Factors by Release Status

PSA Factor	Factor Labels	Overall N(%)	Detained N(%)	Released N(%)	FTA	NCA	NVCA
1. Age at current arrest	<23	758(11.8)	237(10.2)	521(12.7)		X	
	>=23	5,679(88.2)	2,089(89.8)	3,590(87.3)			
2. Current violent offense ⁷	No	4,526(70.3)	1,413(60.8)	3,113(75.7)			X
	Yes	1,911(29.7)	913(39.3)	998(24.3)			
2a. Current violent offense and <= 20 years old	No	6,265(97.3)	2,256(97.0)	4,009(97.5)			X
	Yes	172(2.7)	70(3.0)	102(2.5)			
3. Pending charge at the time of the arrest	No	4,733(73.5)	1,551(66.7)	3,182(77.4)	X	X	X
	Yes	1,704(26.5)	775(33.3)	929(22.6)			
4. Prior misdemeanor conviction	No	2,537(39.4)	662(28.5)	1,875(45.6)		X	
	Yes	3,900(60.6)	1,664(71.5)	2,326(54.4)			

6 VanNostrand, M., & Lowenkamp, C. T. (2013). Assessing pretrial risk without a defendant interview. Houston: Laura and John Arnold Foundation. <https://nicic.gov/assessing-pretrial-risk-without-defendant-interview>

7 Individuals in Pierce County reported that during the historical data period, individuals had multiple required court appearances that led to greater likelihood that many individuals experienced an FTA. There have been changes in policies and practices more recently that have resulted in a reduction in the number of required appearances.

PSA Factor	Factor Labels	Overall N(%)	Detained N(%)	Released N(%)	FTA	NCA	NVCA
5. Prior felony conviction	No	3,489(54.2)	950(40.8)	2,539(61.8)		X	
	Yes	2,948(45.8)	1,376(59.2)	1,572(38.2)			
5a. Prior conviction (misdemeanor or felony)	No	1,683(26.1)	421(18.1)	1,262(30.7)	X		X
	Yes	4,754(73.9)	1,905(81.9)	2,849(69.3)			
6. Prior violent conviction	No	4,493(69.8)	1,362(58.6)	3,131(76.2)		X	X
	Yes, 1 or 2	1,478(23.0)	694(29.8)	784(19.1)			
	Yes, 3 or more	466(7.2)	270(11.6)	196(4.8)			
7. Prior FTA in the past 2 years	No	3,448(53.6)	964(41.4)	2,484(60.4)	X	X	
	Yes, just 1	1,471(22.9)	593(25.5)	878(21.4)			
	Yes, 2 or more	1,518(23.6)	769(33.1)	749(18.2)			
8. Prior FTA older than 2 years	No	2,611(40.6)	710(30.5)	1,901(46.2)	X		
	Yes	3,826(59.4)	1,616(69.5)	2,210(53.8)			
9. Prior sentence to incarceration >= 14 days	No	2,994(46.5)	727(31.3)	2,267(55.1)		X	
	Yes	3,443(53.5)	1,599(68.7)	1,844(44.9)			

Note. For the last three columns, an 'X' indicates which PSA factors are used to predict each of the three pretrial outcomes: FTA = failure to appear; NCA = new criminal arrest; NVCA = new violent criminal arrest.

Table 3 includes important information because understanding the proportion of individuals with each of the factors provides a foundation for the rest of the study. Pretrial assessments involve the accumulation of specific factors such that more factors (generally) equate to higher scores and higher scores (generally) are associated with higher probabilities for someone to experience one of the outcomes.

About 30% (n = 1,911) of the people admitted to jail have a current violent charge and 27% (n = 1,704) have a pending charge at the time of their current booking. In the sample, 74% (n = 4,754) of the people have a conviction for a prior crime, 61% of the people have a prior misdemeanor conviction (n = 3,900), 46% of the people have a prior felony conviction (n = 2,948), and 30% of the people have at least one prior violent conviction (n = 1,944). There are two prior FTA factors included in the PSA scales, 46% (n = 2,989) of the people in the pretrial sample had an FTA in the past 2 years and 59% (n = 3,826) had an FTA older than 2 years.

Table 3 shows that relative to the people in the released group, the people in the detained group have more extensive criminal histories (e.g., prior convictions, prior violent convictions), more people in the detained group are charged with a current violent offense (39% v. 24%, $p < 0.001$), and more have a pending charge at the time of their arrest (33% v. 23%, $p < 0.001$). The descriptive statistics provide a general understanding of the differences between the people in the detained and released groups on the PSA.

Next, we provide a more complete comparison of the detained ($N=2,326$) and released ($N=4,111$) subpopulations and show that they pose a similar likelihood of success for making their court date, not being rearrested, and not being rearrested for a violent crime.

One way to understand the differences between two groups is to consider the average (or mean) score on the scales. Table 4 shows the average scale scores for the detained and released groups. As can be seen, the detained group has higher average scores for each of the scales. However, the differences between detained and released individuals' average scores are moderate. These scale scores differ by 0.74, 0.85, and 0.59 points on the FTA, NCA, and NVCA scales, respectively.

The statistical measure Cohen's d is used to assess whether the magnitude of the differences in the average scale scores between the released and detained groups is meaningful. A common rule of thumb is that Cohen's $d = 0.2$ is a small difference, Cohen's $d = 0.5$ is a moderate difference, and Cohen's $d = 0.8$ is a large difference. The values of Cohen's d for the PSA scores range from 0.48 to 0.57 suggesting that there are moderate differences in the PSA factors scores between detained or released individuals in Pierce County.

Table 4: Average PSA Scores by Release Status

Release Status	Average PSA Score (scores range 1-6)		
	FTA	NCA	NVCA
Detained	3.87	3.84	2.44
Released	3.13	2.99	1.85
Cohen's d	0.48	0.57	0.57

These results are based on historical data and the PSA scores were not used to make release decisions. Therefore, the moderate difference in average scores between released and detained individuals suggests that judges are making intuitive risk calculations and weighing prior convictions, FTAs, and violence as important factors for release decisions.

Table 5: Scale Scores by Released and Detained Status

Score	FTA		NCA		NVCA	
	Detained	Released	Detained	Released	Detained	Released
1	238 (10.2%)	958 (23.3%)	221 (9.5%)	870 (21.2%)	550 (23.7%)	1,895 (46.1%)
2	145 (6.2%)	437 (10.6%)	226 (9.7%)	822 (20.0%)	819 (35.2%)	1,322 (32.2%)
3	557 (24.0%)	1037 (25.2%)	380 (16.3%)	824 (20.0%)	468 (20.1%)	596 (14.5%)
4	478 (22.6%)	745 (18.1%)	748 (32.2%)	936 (22.8%)	371 (16.0%)	233 (5.7%)
5	565 (24.3%)	656 (16.0%)	386 (16.6%)	397 (9.7%)	116 (5.0%)	64 (1.6%)
6	343 (14.8%)	278 (6.8%)	365 (15.7%)	262 (6.4%)	*	*
	2,326	4,111	2,326	4,111	2,324	4,110

* Cell sizes less than 16 are not displayed, as statistics calculated from small cells are unreliable.

The differences in average scale scores provide a good indication of how the released and detained individuals scored overall. Table 5 shows how detained and released individuals were distributed across the scale scores. Overall, individuals with lower PSA scores were more likely to be released and those with higher scores were more likely to be detained.

Validation of the Public Safety Assessment

Now that we have reviewed characteristics of the pretrial sample, we turn to the validation. Validation is a process of using statistical tests to determine if the PSA accurately predicts three pretrial outcomes in Pierce County. The validation uses Pierce County data from January 1, 2017, through December 31, 2018, to see if the PSA has adequate predictive validity and lacks evidence of predictive bias. This means we need to ensure that the PSA factors are strongly associated with each of the outcomes, and there are equal probabilities of outcomes across race and sex.

Validation Sample: 4,111 Released Individuals

The previous discussion focused on the entire pretrial sample of PSA-eligible Pierce bookings (i.e., released and detained individuals charged with felony offenses). For the validation study, the sample needs to be reduced to include only those who were released by removing data for individuals detained for their pretrial period from the sample. We removed the data for the 2,326 individuals detained for their pretrial period, yielding a validation sample of 4,111.

Table 6 shows the PSA scale score distribution for the released PSA-eligible individuals in Pierce County. Most people have low PSA scores in Pierce County, with 59% and 61% of the individuals scoring between 1 and 3 on the FTA and NCA scales and more than 90% scoring between 1 and 3 on the NVCA scale. Relatedly, few people score high, with about 23% (FTA), 16% (NCA), and less than 2% (NVCA) scoring 5 or 6 across the three scales.

Table 6: Scale Score Distribution in Pierce County

Scale Score	FTA		NCA		NVCA	
	N	%	N	%	N	%
1	958	23.30	870	21.16	1,895	46.10
2	437	10.63	822	20.00	1,322	32.16
3	1,037	25.23	824	20.04	596	14.50
4	745	18.12	936	22.77	233	5.67
5	656	15.96	397	9.66	64	1.56
6	278	6.76	262	6.37	1	0.02
Total	4,111		4,111		4,111	

Next, we look at the relationship between the scale scores and each of the pretrial outcomes. For a valid assessment, the outcome rates will increase as scale scores increase such that a greater proportion of people with higher scores experience one of the outcomes. Table 7 shows the numbers of people who were observed to have each of the PSA outcomes by their scale scores. Also shown are the percentages of the individuals with each scale score that experienced the outcome. For example, 23% or 222 of the 958 individuals with an FTA score of 1 (see Table 6) experienced an FTA. If the PSA is valid, we would expect that a greater proportion of individuals would have experienced one of the negative outcomes as the scores increase. With respect to the FTA scores, 23% of those with a score of 1 experienced an FTA compared to 44% of those with a score of 6. There was little variation in the mid-range scores of 2 to 5 with between 31% and 36% of individuals experiencing an FTA.⁷

Table 7: Negative Pretrial Outcomes for Released Individuals by Scale Score

Scale Score	Observed FTA N	% FTA	Observed NCA N	% NCA	Observed NVCA N	% NVCA
1	222	23.17	165	18.97	84	4.43
2	143	32.72	204	24.82	109	8.25
3	378	36.45	229	27.79	86	14.43
4	230	30.87	336	35.90	44	18.88
5	242	36.89	153	38.54	19	29.69
6	123	44.24	118	45.04	0	0
	1,338		1,205		342	

The NCA and NVCA patterns present a more consistent picture of increasing negative outcome rates as the scale scores increase. The NCA rates range from 19% to 45%, and the NVCA rates range between 4% and 30%.

Predictive Validity

The predictive validity of a scale is the degree to which it predicts whether the outcome (e.g., FTA) will occur. We showed that higher PSA scale scores are generally related to the more frequent outcomes, but to provide a more formal statistical analysis we use the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) measure that provides a numerical way to evaluate how well the scales perform.

The AUC is calculated by plotting the correct outcome predictions (i.e., true positives) against the incorrect outcome predictions (i.e., false positives) for each of the scale scores. The AUC values range from 0 to 1, with zero meaning that the assessment is useless because it predicts all successful cases as failures and all cases that fail as successes. Conversely, an AUC of 1 means the assessment flawlessly distinguishes between all people who are successful and those who are unsuccessful on pretrial release. Another thing to consider when looking at the AUCs is that an AUC of 0.5 means the assessment does not distinguish between successful and unsuccessful cases and the assessment is no better than chance. AUCs are not influenced by prevalence rates and, as such, they are a good metric to use when needing to make comparisons among groups and samples. Typically, for pretrial assessment instruments AUCs in the 0.6 to 0.7 range are what are observed.

Table 8 includes the AUCs for the three outcomes in Pierce County. The AUCs for NCA (0.61) and NVCA (0.66) are within the typical range for pretrial assessments. The FTA AUC score is a bit lower but remains statistically significant and above the 0.50 threshold. Future validations in Pierce County are recommended to make sure the predictive validity for FTAs improves if the PSA is implemented.

Table 8: Predictive Validity Results: Area Under the Curve (AUC) Values

	FTA	NCA	NVCA
AUC	0.56	0.61	0.66

Assessing Predictive Bias: Ensuring Equal Probabilities

Predictive bias testing is a process to assess whether the PSA provides equal results for different race and sex groups. This means that, for instance, a Black individual with a score of 2 on the FTA scale will have the same chance of an FTA as a White individual with a score of 2. For any score on the risk scales, there should be equal probabilities of the pretrial outcomes regardless of race or sex. Finding different probabilities of failure within risk scores by race or sex could create ethical and practical challenges related to detention, supervision conditions, and public safety. Assessing predictive bias is complex and challenging. There are many ways to measure predictive bias and because of the mathematical relationships among these tests it is impossible to “pass” on all the measures. This is especially true when base rates are unequal across groups.

The base rates of interest here are the prevalence of each of the three pretrial outcomes in Pierce County. The base rate converts the number or count of each outcome to a proportion of the released population. Therefore, the PSA base rates are the percentages of people who had an FTA, NCA, or NVCA during the study period. In tables 9 and 10, you will see that the overall base rate is 32.5%, 29.3%, and 8.3% for FTA, NCA, and NVCA, respectively, in Pierce County. The overall base rate is interesting because it tells us the nature of the pretrial outcomes in Pierce County, but this does not tell us anything about predictive bias.

For predictive bias testing we need to know whether base rates differ between the subgroups of interest. For example, are there differences in FTA rates between White individuals and people of color or males and females? Looking for base rate differences brings up the important question of: what is a difference? Base rates are unlikely to be identical between groups so statistical tests are used to determine whether observed differences are “statistically significant” or most likely to be “real” and not a result of sampling. To determine significance, we apply a threshold known as a p-value that provides strong confidence that any identified differences are truly differences between the groups. We follow recent practices and set statistical significance levels at $p < 0.001$ because of the large sample sizes in the Pierce County data.⁸ This threshold means that there is less than a 0.1% chance that we would mistakenly suggest there is a difference between groups when there was no difference.

Tables 9 and 10 show the overall and group PSA score base rates for race groups (White and people of color) and sex groups (males and females). There are no significant differences in the base rates by race or groups. FTAs were observed for 32.5% of the sample, NCAs for 29.3%, and NVCA for 8.3%. The 1.3% difference in NVCA rates between White individuals and people of color is small and insignificant. The small differences in the outcome rates by sex and are insignificant.

⁸ Monahan, J., Skeem, J., and Lowenkamp, C. (2017). Age, risk assessment, and sanctioning: Overestimating the old, underestimating the young. *Law and Human Behavior*, 41(2), 191–201.

Table 9: Base Rates for FTA, NCA, and NVCA by Race

Scale	White (N=2,724)	People of color (N=1,287)	Overall (N=4,111)	p-value
FTA	893 (32.8%)	445 (32.1%)	1,338 (32.5%)	0.677
NCA	797 (29.3%)	408 (29.4%)	1,205 (29.3%)	0.945
NVCA	214 (7.9%)	128 (9.2%)	342 (8.3%)	0.148

Table 10: Base Rates for FTA, NCA, and NVCA by Sex

Scale	Male (N=2,952)	Female (N=1,134)	Overall (N=4,111)	p-value
FTA	941 (31.9%)	391 (34.5%)	1,338 (32.5%)	0.121
NCA	897 (30.4%)	304 (26.8%)	1,205 (29.3%)	0.027
NVCA	268 (9.1%)	74 (6.5%)	342 (8.3%)	0.010

Table 11 shows the AUC values are in the fair to good range with the FTA AUC for females 0.532 being the smallest value. As can be seen, none of the p-values approach the 0.001 threshold and, thus, we can conclude that there are no significant differences in the predictive validity of the PSA across race and sex.

Table 11: AUC Values by Race and Sex

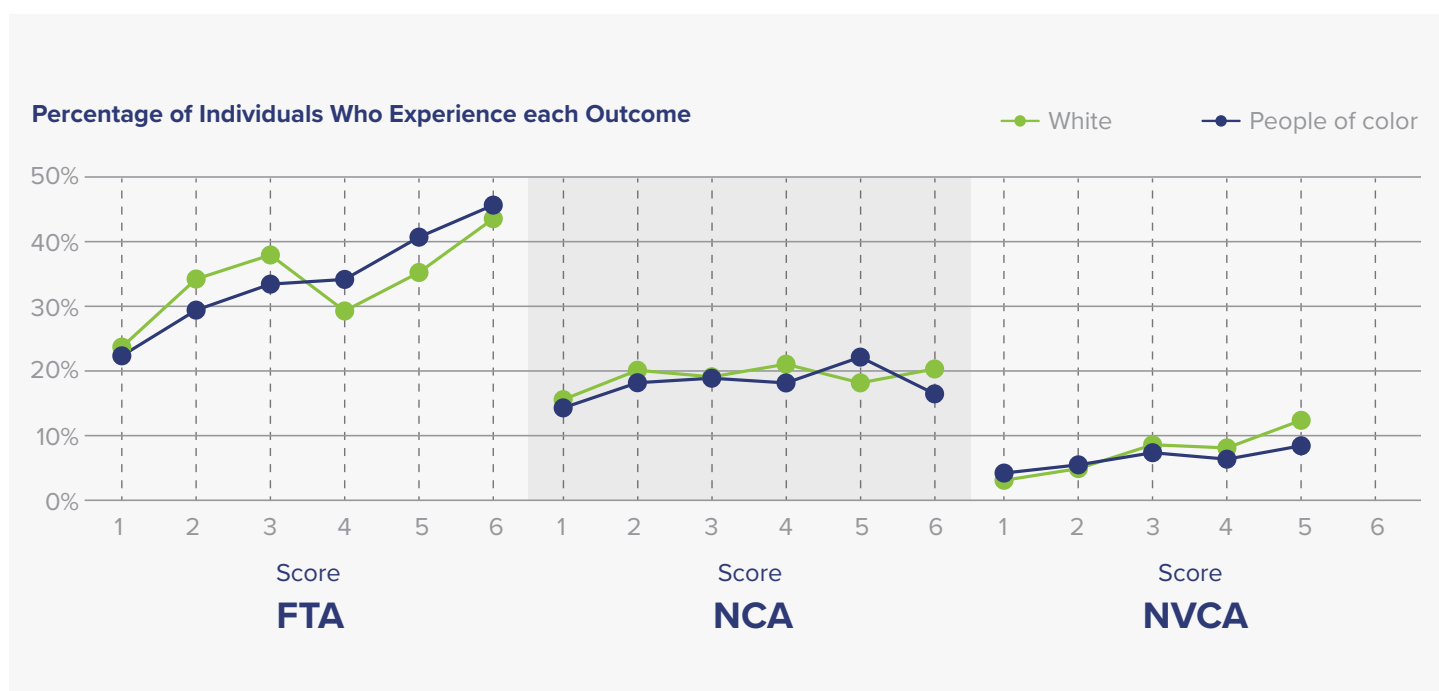
Outcome	People of color	White	p-value
FTA	0.589	0.548	0.035
NCA	0.615	0.600	0.463
NVCA	0.633	0.676	0.176
Outcome	Female	Male	p-value
FTA	0.532	0.572	0.056
NCA	0.621	0.597	0.257
NVCA	0.669	0.653	0.685

Calibration: Equal Rates Across Groups

Calibration examines whether equal proportions of the subgroups have the outcomes of interest. Thus, calibration provides a check for predictive bias by determining whether the PSA scores and outcomes are related to one another in similar ways across, in this case, race and sex—in other words, do we see similar patterns in the PSA scores and outcomes for people of color and White individuals and males and females?

Figure 1 shows the FTA, NCA, and NVCA scores and the proportion of individuals with each score who had the relevant outcome for people of color and white individuals. See the graphs show that the higher PSA scores are associated with higher rates of the outcomes and that there are only small differences between people of color and white individuals. For example, for those with an FTA score of 1, there are about 25% of White individuals and people of color that have an FTA during the pretrial period. White individuals have slightly lower FTA rates for scores of 4 and 5, these increase to nearly 50% for scores of 6 for both white individuals and people of color, but these slight differences are not statistically significant. Importantly, the scores do not consistently differ across several of the scale scores between people of color and whites.

Figure 1: Validation Sample with Negative Outcomes by Race



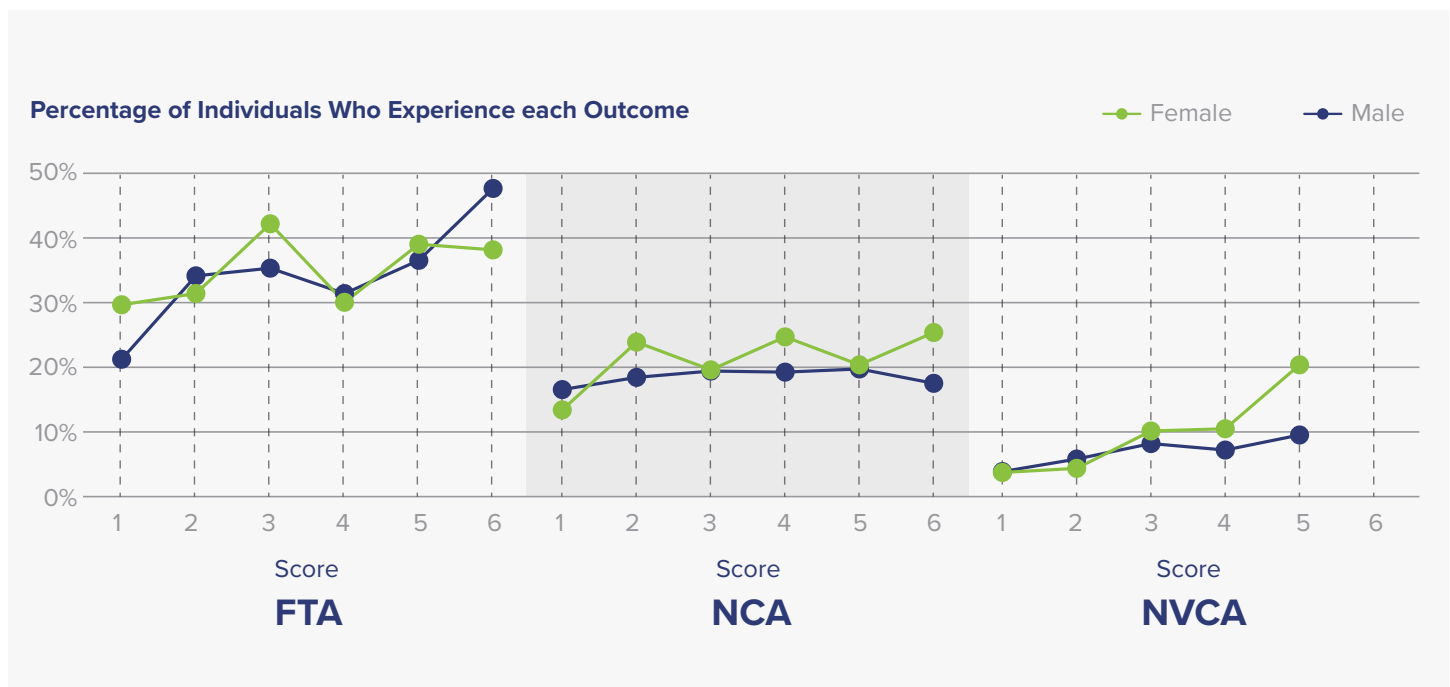
The results are similar for the NCA scales. There is a consistent linear increase in the proportion of individuals experiencing an NCA as scores increase for both race groups. Again, the important issue is that we do not see variation in these outcomes by race – none of the scale scores reach statistical significance.

The NVCA scale demonstrates similar patterns to the NCA scale, but with much lower overall rates. There were too few individuals with scores of 5 or 6 to report. There is parity in the NVCA rates between White individuals and people of color across the scores and linear increases in NVCA rates from about 5% (NVCA = 1) to 20% (NVCA = 4). None of differences between groups are statistically significant.

Although concern about predictive bias in assessments has often focused on differences between White individuals and people of color, there are similar concerns that female individuals may be incorrectly scored higher risk. This concern is known as overprediction, in which one group is scored higher than their true risk level.

Figure 2 provides the same information as in Figure 1, but the groups are males and females. Again, the PSA achieves both criteria of higher scores associated with higher rates of the outcomes and only small differences between males and females. For example, about 21% of males and 29% of females with an FTA score of 1 experienced an FTA. As before, the small differences between males and females are not statistically significant.

Figure 2: Validation Sample with Negative Outcomes by Sex



Similar results are obtained for the NCA scale that shows a generally linear increase with NCAs and scores for males. NCAs increase from about 25% for males with an NCA score of 1 to about 40% for males with a score of 6. The trend for females is similar but less linear. Again, there is no variation in these outcome rates by score between males and females. The NVCA scale demonstrates similar patterns to the NCA scale, but with much lower overall rates. There are too few people to compare males and females for scores above 3. There is parity in failure rates by scores between males and females for NVCA scores of 1-3.

PSA Pretrial Outcomes for Race and Sex Subgroups

So far, we have shown that the PSA provides good classification by race and sex and that there is little difference in the reported outcomes between race and sex groups. In this section, we share results of statistical analyses that were conducted to determine whether the PSA predicts equal probabilities of each of the outcomes for race and sex subgroups. To address this question, we use logistic regression, which is a statistical procedure that estimates the likelihood or probability of an event happening based on a set of factors or variables. Results from the logistic regression analyses indicate whether the PSA scales predict the pretrial outcomes and whether there are strong relationships between risk, race, sex, and the outcomes that would suggest that the PSA scales are biased.

Table 12 provides the results of a set of logistic regression models estimated with the Pierce County data.⁹ The dependent variable for each model is the indicator of whether the event was observed (e.g., for the FTA model, the dependent variable equals 1 if the individual had an FTA or zero if the individual did not have an FTA). Four models are shown for each of the three PSA scales (FTA, NCA, NVCA). Model 1 includes only race (White = 1; people of color= 0) as a covariate—showing the direct relationship between race and the observed outcome. Model 2 includes only the relevant scale score (i.e., FTA score for the FTA model, NCA score for the NCA model, or NVCA score for the NVCA model) as a covariate—showing the direct relationship between the score and the observed outcome. Model 3 includes both the race and scale score—testing the relationship of both variables to the outcome simultaneously. Model 4 includes race, the scale score, and an interaction term (White*Score) that tests whether there is a differential effect of the score by race.

⁹ The values in Tables 12 and 13 are odds ratios which provide an indication of the direction of a relationship. An odds ratio less than 1 means higher values for that variable are associated with less risk (i.e., less likelihood) of the outcome occurring. An odds ratio greater than 1 means higher values for that variable are associated with more risk (i.e., greater likelihood) of the outcome occurring. The p-value indicates whether the odds ratio is significantly different from 1. If the odds ratio isn't significantly different from 1 then there is no relationship between the variable and the outcome. As before, because of the size of the dataset, we are using a p-value of ≤ 0.001 to indicate statistical significance.

Table 12: Logistic Regression Results Testing for Predictive Bias by Race

Variable	Model 1		Model 2		Model 3		Model 4	
	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value
FTA Scale								
White	1.032	0.651			1.011	0.875	1.335	0.077
FTA Score			1.157	<0.001	1.157	<0.001	1.224	<0.001
White * FTA Score							0.917	0.058
Intercept	0.472	<0.001	0.303	<0.001	0.301	<0.001	0.251	<0.001
NCA Scale								
White	0.992	0.916			1.006	0.940	1.088	0.629
NCA Score			1.285	<0.001	1.285	<0.001	1.306	<0.001
White * NCA Score							0.976	0.618
Intercept	0.417	<0.001	0.190	<0.001	0.190	<0.001	0.180	<0.001
NVCA Scale								
White	0.839	0.132			0.935	0.576	0.616	0.073
NVCA Score			1.741	<0.001	1.736	<0.001	1.558	<0.001
White * NVCA Score							1.192	0.088
Intercept	0.102	<0.001	0.029	<0.001	0.030	<0.001	0.039	<0.001

The results across all the models demonstrate insignificant and weak associations between race and the outcomes. Race is included in nine of the twelve models including as an interaction term (model 4) and is insignificant in all models.

Table 12 shows that the PSA scale scores are significant predictors for FTAs, NCAs, and NVCA. The results from the models confirm that increases in scores are associated with statistically significant increases in the likelihood of an FTA, NCA, and an NVCA. Specifically, the odds ratios are 1.224 (FTA), 1.306 (NCA) and 1.558 (NVCA) and these odds ratios are statistically different from 1 as the associated p-values are ≤ 0.001 . These results mean that for each point increase in the FTA, NCA, and NVCA score there is a 22%, 31% and 56% increase in the probability of those outcomes, respectively. Higher scores are related to significantly greater likelihood that someone will miss court or be rearrested during their pretrial release.

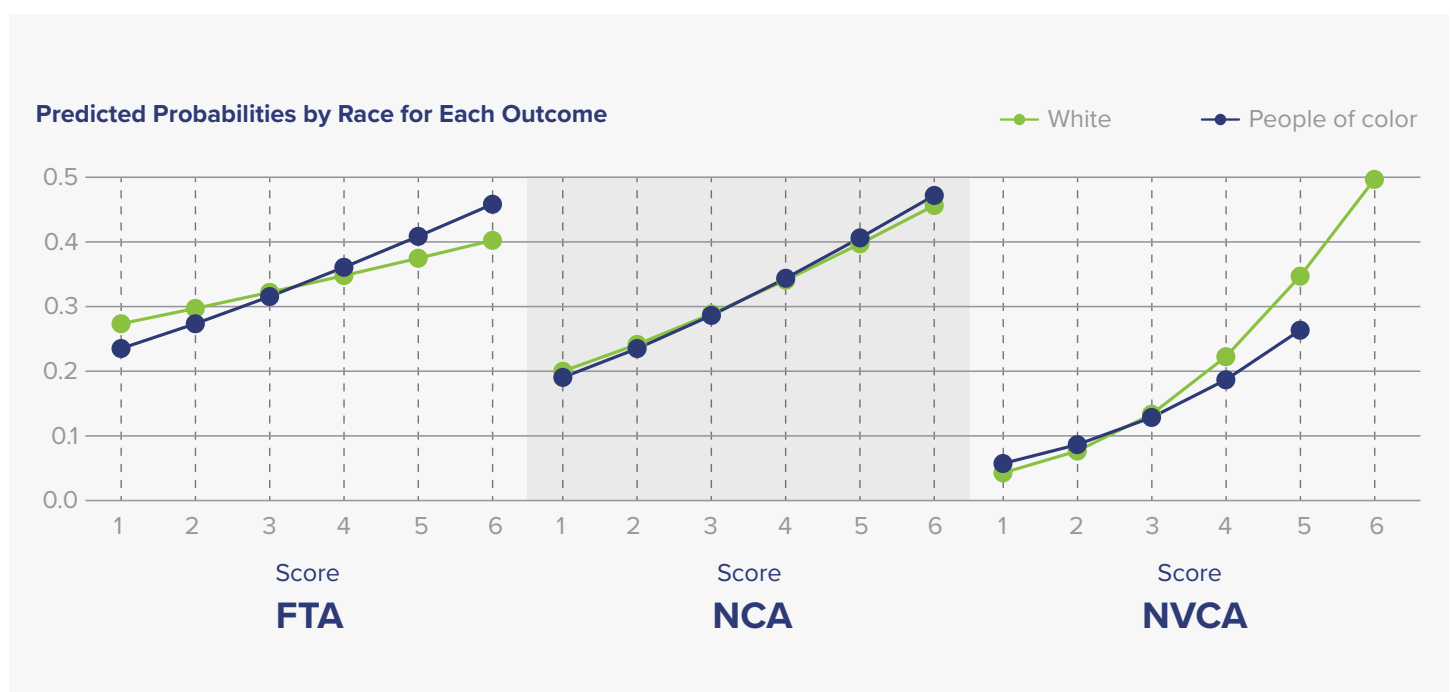
Model 3 tests the relationship between the race variable and the scale score with the outcomes simultaneously. The odds ratio values are similar for each variable in size and direction to those from the simpler models, and provides additional evidence that race is unrelated to the outcomes. The PSA scale scores remain significant, independent predictors of the PSA outcomes.

The final set of models tests whether predicted outcomes for a score are the same for each race group. This is tested by including the interaction term (White*Score) in the models. Specifically, if the resulting odds ratio is significantly different from 1 (i.e., if the p-value is ≤ 0.001) then the results suggest that the scale is providing different results for different racial groups.

Results from Model 4 in Table 12 show that the odds ratios for the interaction term (White*Score) for the three outcomes are not significantly different from 1. In other words, the predictions of the PSA for appearing in court (FTA), being arrest for a new crime (NCA), and experiencing a new violent criminal arrest (NVCA) are not different for White individuals and people of color in Pierce County. The scale scores have the same meaning for White individuals and people of color in Pierce County.

Figure 3 plots the predicted probabilities for the outcomes by race for each score (model 4). The left or y-axis provides the proportion of individuals that are predicted to have each of the outcomes. The blue lines are estimated for people of color and the green lines for White individuals. As with Figures 1 and 2, to assess predictive bias, we are looking at (1) the closeness of the lines since identical predictions would have overlaying lines and highly divergent predictions would have lines with a lot of distance between them; and (2) the general trend in the relationship between the lines as a higher line would indicate higher predicted scores for the subgroup.

Figure 3: Plots of Predicted Probabilities for Pretrial Outcomes by Race



The FTA plot shows slightly higher FTA rates for White individuals relative to people of color for scores of 1 and 2, identical rates for scores of 3 and 4, and slightly higher FTA rates for people of color for scores 5 and 6. Importantly, however, these differences are insignificant (Table 11, model 4).

The NCA plots show nearly identical predicted rates of new arrest across the scale scores for White individuals and people of color. The predicted outcome rates range from about 20% to 45% for scores 1 and 6, respectively, and the small differences are not statistically significant. Similarly, the NVCA plots also show nearly identical predicted outcome rates for White individuals and people of color for scores 1 through 4. For scores of 5 and 6, White individuals have a higher predicted probability of a new violent arrest, but these differences are insignificant (Table 12, model 4).

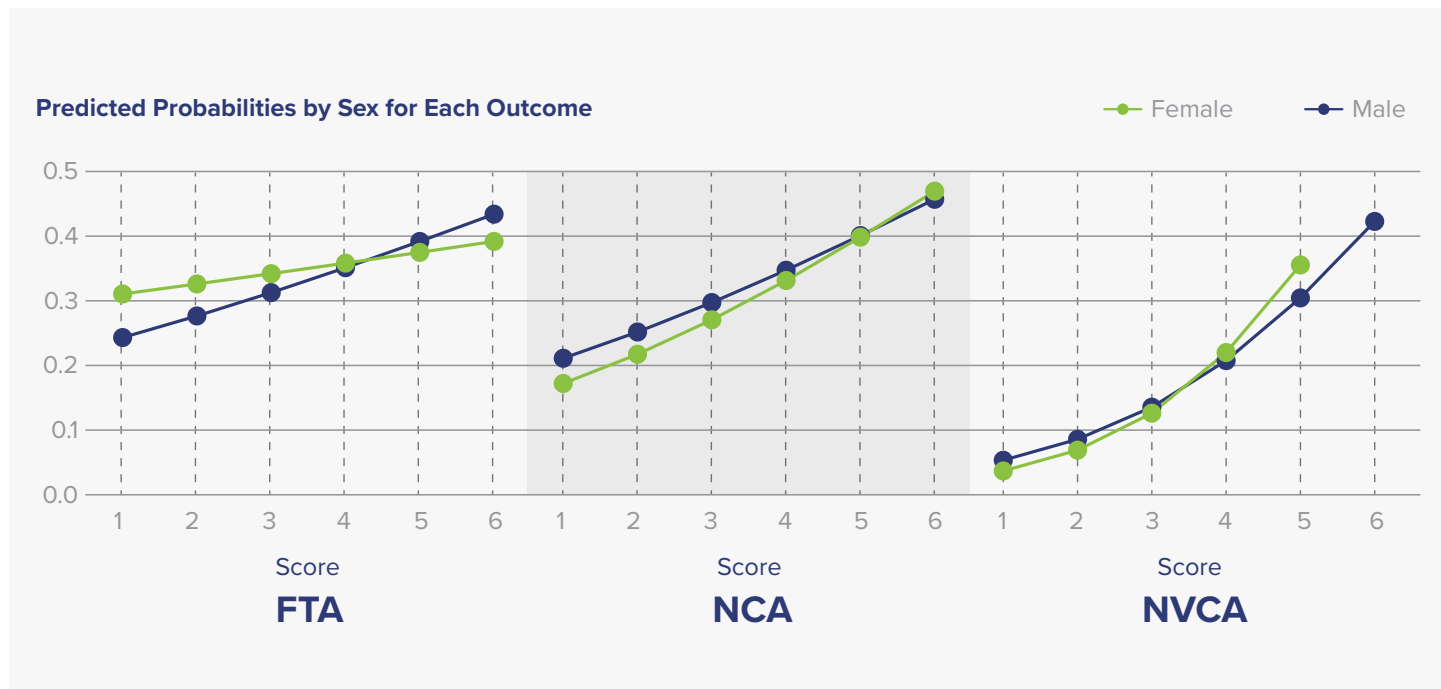
Table 13 provides the logistic regression model results examining the relationship between sex and the outcomes. Sex is included in nine of the twelve models including as an interaction term (model 4) and is insignificant in all models. Model 3 tests the relationship between the sex variable and the scale score with the outcomes simultaneously. The odds ratio values are similar for each variable in size and direction to those from the simpler models, which provides additional evidence that sex is not related to the outcomes, but the PSA scale scores are significant, independent predictors of the PSA outcomes. Results from Model 4 show that the odds ratios for the interaction term (Male *Score) for the three outcomes are insignificant. In other words, the predictions of the PSA for appearing in court (FTA), being arrest for a new crime (NCA), and experiencing a new violent criminal arrest (NVCA) are not different for Males and Females.

Table 13: Logistic Regression Results Testing for Predictive Bias by Sex

Variable	Model 1		Model 2		Model 3		Model 4	
	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value	Odds Ratio	p-value
FTA Scale								
Female	0.889	0.112			0.899	0.155	0.643	0.009
FTA Score			1.157	<0.001	1.152	<0.001	1.074	0.061
Female * FTA Score							1.108	0.027
Intercept	0.526	<0.001	0.303	<0.001	0.332	<0.001	0.418	<0.001
NCA Scale								
Female	1.192	0.023			1.132	0.118	1.369	0.090
NCA Score			1.285	<0.001	1.278	<0.001	1.337	<0.001
Female * NCA Score							0.941	0.253
Intercept	0.366	<0.001	0.190	<0.001	0.178	<0.001	0.155	<0.001
NVCA Scale								
Female	1.430	0.009			1.227	0.141	1.736	0.071
NVCA Score			1.741	<0.001	1.733	<0.001	1.958	<0.001
Female * NVCA Score							0.855	0.194
Intercept	0.070	<0.001	0.029	<0.001	0.025	<0.001	0.019	<0.001

Figure 4 plots the predicted probabilities for the outcomes by sex for each score (model 4). The left or y-axis provides the proportion of individuals that are predicted to have each of the outcomes. The green lines are estimated for female and the blue lines for male individuals. The FTA plots show little difference in the predicted probabilities between males and females for FTAs by scale score. The slight differences between scores are not statistically significant. Results are similar for the FTA analysis—small differences are not statistically significant. Third, the NVCA plots show that the PSA provides nearly identical rates for an NVCA for males and females for scores of 1 and 2. Male have higher predicted outcome rates for scores of 3 to 5, but these differences are insignificant.

Figure 4: Plots of Predicted Probabilities for Pretrial Outcomes by Sex



Conclusion

Our analyses assess the predictive validity of the PSA when applied to an historical jail booking cohort in Pierce County. Predictive validity measures how well the PSA classifies individuals in Pierce County such that we would expect individuals receiving lower PSA scores to have lower rates of the negative pretrial outcomes and higher rates of court attendance, and lower rates of rearrests for any charge or a violent charge. Validations are designed to answer the basic question: does the PSA measure what it is intended to measure? The findings reveal the PSA is a good predictive measure of the likelihood of pretrial success in Pierce County.

There are important concerns about bias by race and sex when making pretrial release decisions. For this reason, we conducted a series of statistical tests to assess whether the PSA shows the presence of statistical bias to determine whether the PSA performs similarly no matter one's race or sex. These analyses address whether a given score on the PSA has the same probability of success regardless of race or sex. For instance, do individuals of color and white individuals with a score of 6 have similar likelihoods of success? The findings reveal that the PSA makes the same predictions for pretrial outcomes regardless of the race or sex of individuals in Pierce County.

The results presented here suggest that the PSA scores of the historical Pierce County data yielded little evidence of predictive bias and no evidence of bias that negatively impacts people of color or females.

Pierce County Draft Historical Validation File

Filtering criteria

The following bookings are omitted from PSA factor eligibility:

Post-trial bookings:	1,407
Probation/parole only:	3,513
Process-related charges ⁱ :	1,080
Juvenile defendants:	41
Sub-total of filtered bookings:	6,041

Bookings are omitted due to the following:

Subsequent booking in 2017-2018 ⁱⁱ :	10,631
Sub-total of Filtered bookings:	10,631

Bookings omitted from validation analysis:

No court data ⁱⁱⁱ :	177
No valid court data ^{iv} :	1,593
Jail release reason indicates no charges were filed ^v :	1,822
Misdemeanor, Municipal Court, or District Court cases:	9,352
Sub-total of Filtered bookings:	12,944

Bookings omitted from outcome analysis:

Detained pretrial:	2,326
Sub-total of Filtered bookings:	2,326

Date range for cohort: All bookings from 2017-2018

Additional documentation

ⁱ Bookings that only have process-related charges associated with the booking refer to out-of-county warrants/holds, correctional transfers, ICE detainees, or transfers to programs (e.g., Drug Court).

ⁱⁱ For individuals booked pretrial during the date range of the validation study, only their first pretrial booking is included in the study cohort. Subsequent (second or higher-order) bookings that occur during the validation study date range are omitted from the outcome analysis. This means that the unit of analysis for the validation study is individuals, not bookings.

ⁱⁱⁱ No court data were found for the people associated with these bookings.

^{iv} These bookings do not have a court record with a disposition date on or after the jail booking date and a related disposed charge. Without a court disposition date, it was not possible to define the pretrial period or determine pretrial release status. Pretrial status for these bookings was determined by jail release reason (e.g., bail bond, cash bail, PR).

^v These booking resulted in release from jail due to no charges being filed. Because no charges were filed, these bookings do not have a court case and were resolved at jail release. For this reason, the bookings are omitted from the outcome analysis.

