ADVANCING PRETRIAL
POLICY AND RESEARCH

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

Matthew DeMichele, RTI International
Stephen Tueller, RTI International
Kimberly Janda, RTI International
Debbie Dawes, RTI International
Pamela K. Lattimore, RTI International

JANUARY 2023

Table of Contents

Executive Summary **4**

Introduction **5**

What is a Pretrial Assessment? **5**

Why is it important to Validate Pretrial Assessments? **5**

Who is booked into Thurston County Jail? **6**

How do PSA-eligible individuals score on the PSA factors? **8**

Validation of the Public Safety Assessment **11**

Validation Sample: 3,066 Released Individuals **11**

Predictive Validity **13**

Assessing Predictive Bias: Ensuring Equal Probabilities **14**

Calibration: Equal Rates Across Groups **16**

PSA Pretrial Outcomes for Race and Sex Subgroups **18**

Conclusion **23**

Exhibits

Figure 1: Validation Sample with Negative Outcomes by Race 16

Figure 2: Validation Sample with Negative Outcomes by Sex 17

Figure 3: Plots of Predicted Probabilities for Pretrial Outcomes by Race 20

Figure 4: Plots of Predicted Probabilities for Pretrial Outcomes by Sex 22

Table 1: Descriptive Characteristics by Release Status 6

Table 2: Most Serious Charge for Individuals Booked into the Thurston County Jail (2017-2018) 7

Table 3: PSA Factors by Detained and Release Status 8

Table 4: Average PSA Scores by Release Status 9

Table 5: Scales Scores by Released and Detained Status 10

Table 6: Scale Score Distribution in Thurston County 11

Table 7: Negative Pretrial Outcomes for Released Individuals by Scale Score 12

Table 8: Predictive Validity Results: Area Under the Curve (AUC) Values 13

Table 9: Base Rates for FTA, NCA, and NVCA by Race 15

Table 10: Base Rates for FTA, NCA, and NVCA by Sex 15

Table 11: AUC Values by Race and Sex 15

Table 12: Logistic Regression Results Testing for Predictive Bias by Race 19

Table 13: Logistic Regression Results Testing for Predictive Bias by Sex 21

Acknowledgments

The authors would like to thank the individuals in Thurston County who were willing to share data, their time, and expertise to support RTI's research on the Public Safety Assessment. We especially thank Ray Brady, Marianne Clear, Andrew Peterson, and John Snaza . 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 the support of Megan Comfort, Monica Shepherd, and other members of RTI's Racial Community and Justice Committee. 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

This report presents findings from a validation and predictive bias test of the Public Safety Assessment (PSA) in Thurston County, Washington. Thurston 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 the 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, Thurston County officials are engaged in efforts across their local criminal legal system to develop improvements to their pretrial system. These improvements are intended to improve pretrial decision making, reduce pretrial detention, and reduce racial disparities. The PSA is an assessment tool that holds promise for facilitating such advancements and has been associated with reductions in missed court appearances and new crimes, and fewer admissions.[1] RTI is contributing to Thurston County's 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 fail to appear in court (FTA), be arrested for a new crime (NCA), and be arrested for a new violent crime (NVCA) if they are released pretrial. The PSA consists of 9 factors used across three scales to predict FTA, NCA, and NVCA. The three PSA scores (FTA, NCA, NVCA) range from 1- 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 from adults booked into the Thurston County jail on a new charge between January 1, 2017, and December 31, 2018. On average, individuals booked in Thurston County score about 3 on the FTA and NCA and 2 on the NVCA scales, with nearly 60% of the pretrial sample scoring between 1-2 on the FTA and NCA scale and nearly 80% scoring 1-2 on the NVCA scale.

One concern often expressed about pretrial release assessment instruments is that people of color and females will be scored too high relative to their actual likelihood of an FTA, NCA or NVCA. The concern under those circumstances is that if conditions of release are predicated on risk scores, people of color or females may be disadvantaged.

The results reported here show that in Thurston County, the PSA factor scores are associated with pretrial outcomes. Further, the bias analyses do not provide evidence that the PSA scores do not exhibit predictive bias related to race and sex. Specifically, there is no evidence that people of color or women are being scored higher than their actual outcome rates (i.e., no overprediction). These results were based on an historical cohort. Proper use of the PSA will require ongoing research to ensure that the PSA remains valid for Thurston 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.

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 Thurston County, Washington. The validation uses historic data to examine how the PSA would have performed in a recent historical period to determine whether the PSA is valid for implementation in Thurston 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 risk (or likelihood) that an individual with specific characteristics will be charged with a new crime and will miss a court appointment if they are released from jail pretrial. 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 stakeholders 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 stakeholders, and does not exhibit predictive bias.

Who is booked into Thurston County Jail?

The data for our study include 4,490 jail bookings that occurred in Thurston County, Washington, between January 1, 2017, and December 31, 2018, for new cases assigned to the Municipal, District, and Superior Courts. Table 1 shows the race and ethnicity of individuals booked into Thurston 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 Thurston County, released individuals were detained in jail an average (median) of 2 days and detained individuals spent an average (median) of 51 days in jail. About one-third (n = 1,424, 32%) of the booked sample was detained pretrial and about two-thirds were released (n = 3,066, 68%).

Table 1: Descriptive Characteristics by Release Status

	Detained (N=1,424)	Released (N=3,066)	Overall (N=4,490)	2019 Census (N=290,536)§
Individuals of Color	364(25.60%)	716(23.40%)	1,080(24.00%)	25.90%
Asian	93(6.50%)	175(5.70%)	268(6.00%)	6.30%
American Indian/Alaska Native	NA†	NA†	NA†	1.80%
Black	159(11.20%)	287(9.40%)	446(9.90%)	3.60%
Hawaiian/Pacific Islander	27(1.90%)	50(1.60%)	77(1.70%)	1.00%
Hispanic/Latin/Mexican	85(6.00%)	204(6.70%)	289(6.40%)	9.40%*
Two or More Races	NA†	NA†	NA†	5.80%
Unknown	34(2.40%)	88(2.90%)	122(2.70%)	NA‡
White	1,026(72.10%)	2,262(73.80%)	3,288(73.20%)	74.10%

*For the US Census, Hispanics may be of any race and are included in applicable race categories. This may cause the 2019 Census column to sum to greater than 100%

†This category is not defined in the Thurston County jail 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 individuals in pretrial system are all over 18.

About three-quarters of the cohort are White individuals (73%), and the average (median) age is 33 years old (min=18, max=86). About one-tenth of the cohort are Black (n = 446), 8% (n = 268) are Asian, Hawaiian, or Pacific Islander, and 6% (n = 289) are Hispanic/Latin/Mexican. 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 9.9% of the booked population and 9.4% of the released population. In Table 1, the booking rate for Black individuals is higher than their presence in the population (3.6%). There is a lack of evidence of disproportionate incarceration in the Thurston County jail with individuals of color making up about 26% of the general population and 24% of the booked population.

Table 2 provides information on the most serious charge for the people booked into the Thurston County jail between January 1, 2017, and December 31, 2018. Information is provided for all 4,490 PSA-eligible individuals (both released and detained) and the 3,066 who were released. In order of prevalence, the most serious charge associated with an admission was classified as a violent (33%), property (29%), public order (26%), drug (12%), and other (0%) charge.[2]

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

Most serious charge overall category	Total Booked N (percentage)	Released N (percentage)
Violent Offenses	1,479(32.90%)	933(30.40%)
Property Offenses	1,301(29.00%)	854(27.90%)
Drug Offenses	535(11.90%)	384(12.50%)
Public Order Offenses	1,170(26.10%)	891(29.10%)
Other Offenses	5(0.11%)	4(0.130%)
Total	**4,490(100.00%)**	**3,066(100.00%)**

2 Charge categories are derived from the National Corrections Reporting Program broad charge categories using the most serious offense charged.

How do PSA-eligible individuals score on the PSA factors?

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 Thurston 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. 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 someone experiencing an FTA, NCA, or NVCA.

Table 3: PSA Factors by Detained and Release Status

PSA Factor	Factor Labels	Overall N(%)	Detained N(%)	Released N(%)	FTA	NCA	NVCA
1. Age at current arrest	<23	499(11.1)	139(9.8)	360(11.7)		X	
	>=23	3,991(88.9)	1,285(90.2)	2,706(88.3)			
2. Current violent offense	No	3,283(73.1)	975(68.5)	2,308(75.7)			X
	Yes	1,207(26.9)	449(31.5)	758(24.7)			
2a. Current violent offense and <= 20 years old	No	4,389(97.8)	1,396(98.0)	2,993(97.6)			X
	Yes	101(2.2)	28(2.0)	73(2.4)			
3. Pending charge at the time of the arrest	No	2,804(62.4)	819(57.5)	1,985(64.7)	X	X	X
	Yes	1,686(37.6)	605(42.5)	1,081(35.3)			
4. Prior misdemeanor conviction	No	1,515(33.7)	358(25.1)	1,157(37.7)		X	
	Yes	2,975(66.3)	1,066(74.9)	1,909(62.3)			
5. Prior felony conviction	No	2,546(56.4)	660(46.3)	1,886(61.5)		X	
	Yes	1,944(43.4)	764(53.7)	1,180(38.5)			
5a. Prior conviction (misdemeanor or felony)	No	1,039(23.1)	250(17.6)	789(25.7)	X		X
	Yes	3,451(76.9)	1,174(82.4)	2,277(74.3)			
6. Prior violent conviction	No	3,225(71.8)	888(62.4)	2,337(76.2)		X	X
	Yes, 1 or 2	687(15.3)	305(21.4)	382(12.5)			
	Yes, 3 or more	578(12.9)	231(16.2)	347(11.3)			

PSA Factor	Factor Labels	Overall N(%)	Detained N(%)	Released N(%)	FTA	NCA	NVCA
7. Prior FTA in the past 2 years	No	2,199(49.0)	596(41.9)	1,603(52.3)	X	X	
	Yes, just 1	960(29.6)	297(20.9)	663(21.6)			
	Yes, 2 or more	1,331(21.4)	531(37.3)	800(26.1)			
8. Prior FTA older than 2 years	No	1,802(40.1)	475(33.4)	1,327(43.3)	X		
	Yes	2,688(59.9)	949(66.6)	1,739(56.7)			
9. Prior sentence to incarceration >= 14 days	No	2,121(47.2)	514(36.1)	1,607(52.4)		X	
	Yes	2,369(52.8)	910(63.9)	1,459(47.6)			

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 are about the accumulation of specific factors such that more factors (generally) equate to higher scores and higher scores (generally) equate to higher probabilities for someone to experience one of the outcomes.

About 27% (n = 1,207) of the booked individuals have a current violent charge and 38% (n = 1,686) have a pending charge at the time of their current booking. The sample includes 77% (n = 3,451) with a conviction for a prior crime, 66% have a prior misdemeanor conviction (n = 2,975), 43% have a prior felony conviction (n = 1,944), and 28% have at least one prior violent conviction (n = 1,265). There are two factors that measure prior FTA included in the PSA scales, 51% (n = 2,291) and 60% (n = 2,688) of the people in the pretrial sample had an FTA in the past 2 years and an FTA older than 2 years, respectively.

Table 3 shows that relative to the released group, the detained group has more extensive criminal histories (e.g., prior convictions, prior violent convictions), more are charged with a current violent offense (32% v. 25%), and more have a pending charge at the time of their arrest (43% v. 35%). The descriptive statistics provide a general understanding of the differences between the detained and released populations on the PSA.

Next, we provide a more complete comparison of the detained (N=1,424) and released (N=3,066) subpopulations and show that the detained and released populations 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 the average scores for the detained and released groups are moderate. The average scale scores differ by 0.46, 0.57, and 0.38 of 1 point 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 0.28 to 0.36 suggesting that there are small differences in the PSA scores between detained or released individuals in Thurston County. The small Cohen's d (= 0.36) means that if we were to randomly compare detained

individuals to released individuals, we would find that about 60% of the detained have higher scores than the released (i.e., probability of superiority).

Table 4: Average PSA Scores by Release Status

Release Status	Average PSA Score (scores range 1-6)		
	FTA	NCA	NVCA
Detained	3.95	3.90	2.34
Released	3.49	3.33	1.96
Cohen's d	0.28	0.36	0.36

These results are based on historical data and the PSA scores were not used to make release decisions. Therefore, the higher average scale scores for detained individuals compared to released individuals suggest that judges were making intuitive risk calculations that result in detaining people with higher criminal history scores.

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.

Table 5: Scale Scores by Released and Detained Status

Score	FTA		NCA		NVCA	
	Detained	Released	Detained	Released	Detained	Released
1	158	555	153	541	355	1,203
2	108	351	122	526	550	1,106
3	268	601	239	542	279	502
4	269	577	385	691	169	181
5	362	576	254	381	67	70
6	259	406	271	385	*	*

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 Thurston County. The validation uses Thurston 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: 3,066 Released Individuals

The previous discussion focused on the entire pretrial sample of PSA-eligible Thurston bookings (i.e., individuals who were released and detained). For the validation study, the sample needs to be reduced to those who were released by removing individuals detained for their pretrial period from the sample. We removed the data for the 1,424 individuals detained for their pretrial period yielding a validation sample of 3,066.

Table 6 shows the PSA scale score distribution for the released PSA-eligible individuals in Thurston County. About half of the individuals scored between 1 and 3 and half scored between 4 and 6 on the FTA and NCA scales. More than 90% scored between 1 and 3 on the NVCA scale and less than 3% were scored either 5 or 6.

Table 6: Scale Score Distribution in Thurston County

Scale Score	FTA		NCA		NVCA	
	N	%	N	%	N	%
1	555	18.10	541	17.65	1,203	39.24
2	351	11.45	526	17.16	1,106	36.07
3	601	19.60	542	17.68	502	16.37
4	577	18.82	691	22.54	181	5.90
5	576	18.79	381	12.43	70	2.28
6	406	13.24	385	12.56	4	0.13

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 individuals 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, 57 individuals with an FTA score of 1 had an FTA of the 555 individuals (see Table 6) who had an FTA score of 1 or 10.27% of that group. If the PSA is valid, we would expect that a greater proportion of individuals would have experienced the negative outcomes as the scores increase. For FTAs, the proportion of individuals who experienced an FTA increases from 10% to 40% as the FTA scores increase from 1 to 6. The base rate (or overall average) for FTAs is 24.5%, which is similar to the rate for people with a score 3 or 4 on the FTA scale.

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

Scale Score	Observed FTA N	% FTA	NCA N	% NCA	NVCA N	% NVCA
1	57	10.27	32	5.91	31	2.58
2	40	11.40	38	7.22	36	3.25
3	112	18.64	71	13.1	38	7.57
4	170	29.46	151	21.85	11	6.08
5	208	36.11	98	25.72	13	18.57
6	163	40.15	116	30.13	0	0
Total	**750**	**24.46**	**506**	**16.50**	**129**	**4.21**

Similar patterns are observed for the NCA and NVCA outcomes. Negative outcome rates increase as the scale scores increase. The NCA rates range from 6% to 30% with a 16.5% base rate, and the NVCA rates range between 3% and 19% with a 4.2% base rate.

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 uniform 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 positive) for each of the scale scores. The AUC values range from 0 to 1, with 0 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 people who are successful and those who are unsuccessful on pretrial release. An AUC of 0.5 means 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 making comparisons across groups and samples. Typically, for pretrial assessment instruments AUCs in the 0.6 to 0.70 range are what are observed.

Table 8 shows the AUCs for the three PSA outcomes for the Thurston County historical data. As can be seen in Table 8, the PSA yields AUCS in this range (0.67 for the FTA scale, 0.68 for the NCA scale, and 0.64 for the NVCA scale). Thus, AUC values for each predictive validity analysis is in what is considered the “Good” (0.64 to 0.70) range.

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

	FTA	NCA	NVCA
AUC	0.67	0.68	0.64

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, which we describe in detail later.

The base rates of interest here are the prevalence of each of the three pretrial outcomes in Thurston 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. 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? Base rates are unlikely to be identical 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 Thurston County data. 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). The overall base rates are 24.5%, 16.5%, and 4.2% for FTA, NCA, and NVCA, respectively. The overall base rates describe the pretrial outcomes in Thurston County. More importantly, there are no significant differences in the base rates by race or sex groups. The 2.3% difference in NCA rates and the 1.1% difference in NVCA rates between White individuals and people of color is small and insignificant (Table 9). The FTA rand NCA rates are nearly identical across racial categories. There are also some small insignificant differences in the outcome rates by sex (Table 10).

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

Scale	White (N=2,262)	People of Color (N=804)	Overall (N=3,066)	p-value
FTA	559 (24.7%)	191 (23.8%)	750 (24.5%)	0.621
NCA	360 (15.9%)	146 (18.2%)	506 (16.5%)	0.157
NVCA	89 (3.9%)	40 (5.0%)	129 (4.2%)	0.246

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

Scale	Male (N=2,168)	Female (N=890)	Overall (N=3,066)	p-value
FTA	521 (24.0%)	228 (25.6%)	750 (24.5%)	0.379
NCA	365 (16.8%)	140 (15.7%)	506 (16.5%)	0.488
NVCA	99 (4.6%)	30 (3.4%)	129 (4.2%)	0.163

Table 11 shows the AUC values are in the fair to good range, with the NVCA AUC for females of 0.57 being the smallest value. As can be seen, none of the p-values approach the 0.001 threshold and, thus, we can conclude that the PSA has similar validity scores by race or sex and 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

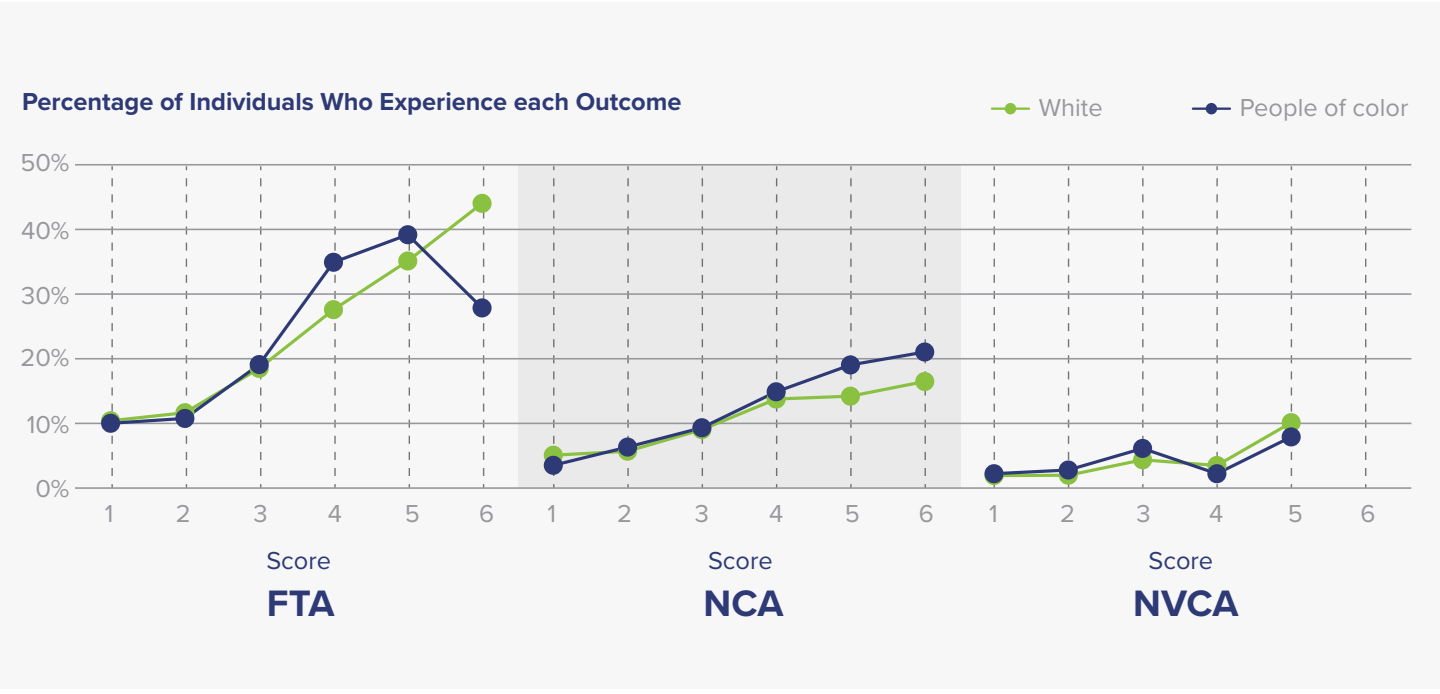
Outcome	People of color	White	P-value
FTA	0.65	0.68	0.380
NCA	0.71	0.66	0.063
NVCA	0.64	0.64	0.922
Outcome	**Female**	**Male**	**P-value**
FTA	0.64	0.68	0.045
NCA	0.65	0.69	0.206
NVCA	0.57	0.66	0.170

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 our 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. 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 10% of White individuals and people of color who had an FTA during the pretrial period. White individuals have slightly lower FTA rates for scores of 4 and 5, and somewhat higher rates for scores of 6 but these differences are not statistically significant (p=0.007). Importantly, the scores do not consistently differ across the scale scores between people of color and White individuals.

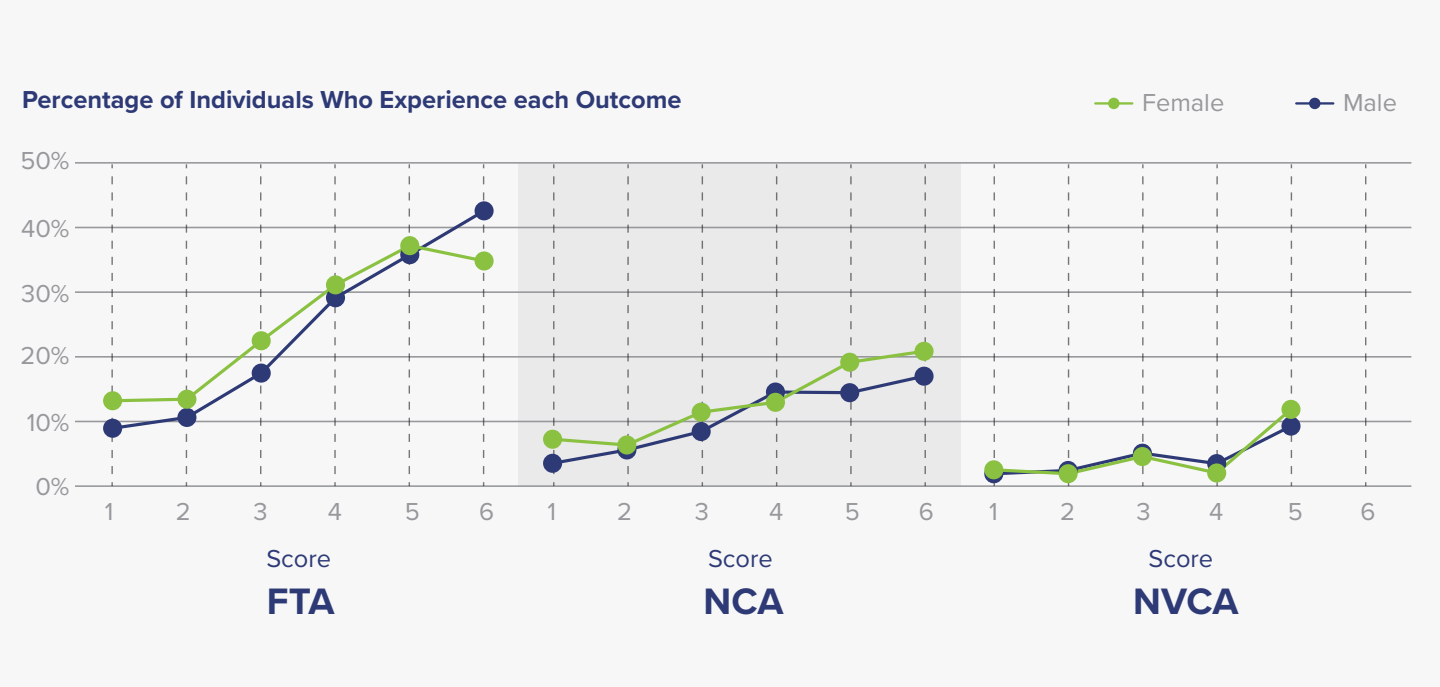
Figure 1: Validation Sample with Negative Outcomes by Race



The results are similar for the NCA scale. There is a consistent linear increase with the proportion of individuals who experience an NCAs with increases in the NCA score across both race groups. Again, the important issue is that there is not variation in these outcomes by race.

Although concern about predictive bias in assessments often focuses 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 9% of Males and 13% of females scoring 1 have an FTA. The small differences in outcomes across the FTA scale scores are not statistically significant (p = 0.178).

Figure 2: Validation Sample with Negative Outcomes by Sex



Similar results are obtained for the new arrest scale that shows a consistent linear increase with NCAs and scores for males and females. Again, there is no variation in outcome rates by score between males and females.

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 Thurston County data.[3] The dependent variable for each logistic regression 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.

3 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	1.053	0.588			0.996	0.972	0.758	0.298
FTA Score			1.475	<0.001	1.475	<0.001	1.403	<0.001
White * FTA Score							1.072	0.270
Intercept	0.312	<0.001	0.075	<0.001	0.076	<0.001	0.092	<0.001
NCA Scale								
White	0.853	0.141			0.823	0.081	1.284	0.427
NCA Score			1.500	<0.001	1.501	<0.001	1.625	<0.001
White * NCA Score							0.895	0.128
Intercept	0.222	<0.001	0.045	<0.001	0.051	<0.001	0.037	<0.001
NVCA Scale								
White	0.782	0.208			0.770	0.183	0.706	0.442
NVCA Score			1.608	<0.001	1.612	<0.001	1.571	0.002
White * NVCA Score							1.036	0.832
Intercept	0.052	<0.001	0.015	<0.001	0.018	<0.001	0.020	<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 NVCAs. The results from model 2 confirm that increases in scores are associated with statistically significant increases in the likelihood of an FTA, NCA, and NVCA. Specifically, the odds ratios are 1.48 (FTA), 1.50 (NCA), and 1.61 (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 48%, 50%, and 61% 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.

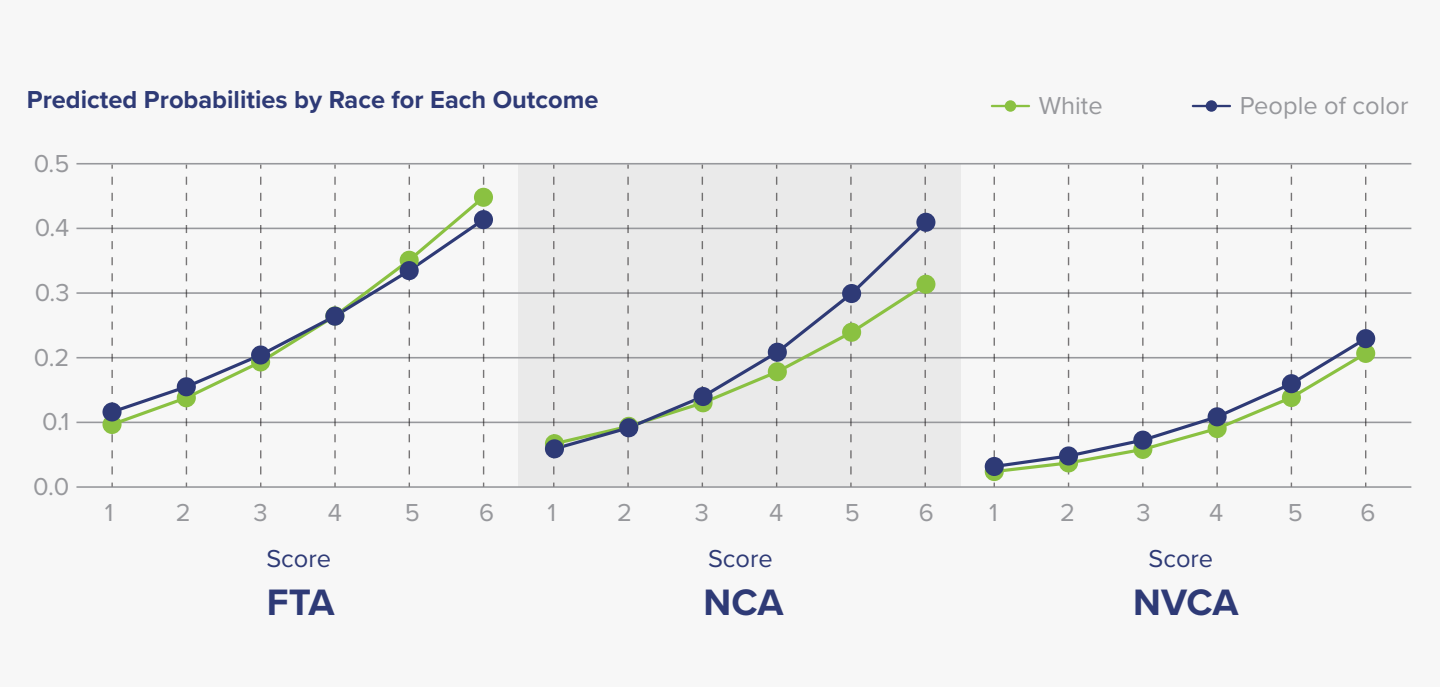
Model 3 tests the relationship between race and the scale score with the outcomes simultaneously. The odds ratios are similar for each variable in size and direction to those from the simpler models, and provides additional evidence that race is unrelated to the outcomes. The PSA scale scores remain 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.[4] 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 Thurston County. The scale scores have the same meaning for White individuals and people of color in Thurston 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 plots show slightly higher FTA rates for people of color relative to White individuals for scores of 1 to 3, identical rates for scores of 4 and 5, and slightly lower FTA rates for people of color for score 6. These differences are not statistically significant (table 12, model 4).

The NCA plots show similar, insignificant differences between the two lines. Predicted outcome rates for White individuals and people of color are nearly identical for scores of 1 to 3. White individuals have slightly lower predicted NCA rates for scores of 4 to 6, but, again, these differences are insignificant. Similarly, the NVCA plots show that people of color have a higher predicted probability of a new violent arrest, but these differences are insignificant.

4 This type of analysis is referred to as moderation analysis.

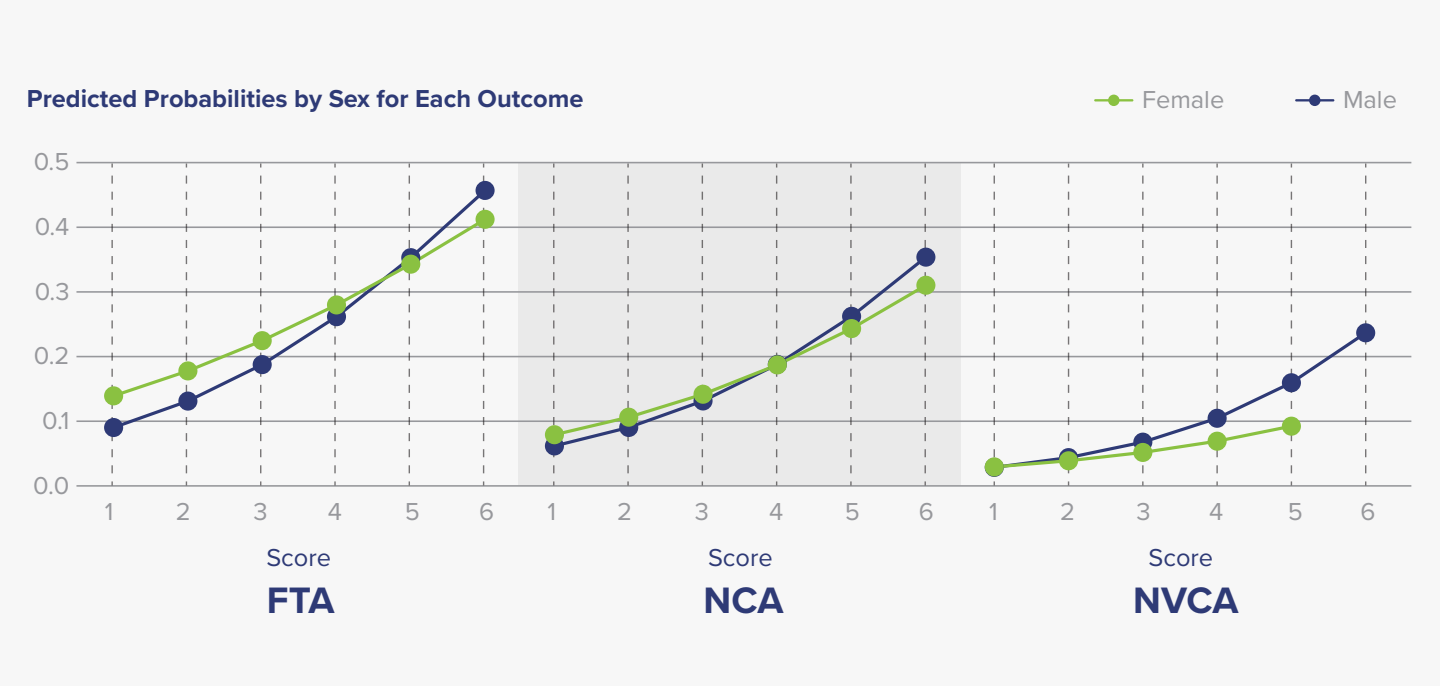
Table 13 provides the logistic regression model results examining the relationship between sex and the outcomes. Sex is included in nine of the twelve models including as an interaction term (model 4) and is insignificant in all models. Model 3 tests the relationship between the sex variable and the 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 failing to appear 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	0.918	0.354			0.907	0.306	0.525	0.011
FTA Score			1.475	<0.001	1.475	<0.001	1.346	<0.001
Female * FTA Score							1.148	0.021
Intercept	0.344	<0.001	0.075	<0.001	0.081	<0.001	0.116	<0.001
NCA Scale								
Female	1.085	0.455			0.995	0.964	0.684	0.201
NCA Score			1.500	<0.001	1.501	<0.001	1.404	<0.001
Female * NCA Score							1.101	0.176
Intercept	0.187	<0.001	0.045	<0.001	0.045	<0.001	0.058	<0.001
NVCA Scale								
Female	1.372	0.137			1.195	0.408	0.819	0.675
NVCA Score			1.608	<0.001	1.594	<0.001	1.393	0.058
Female * NVCA Score							1.184	0.386
Intercept	0.035	<0.001	0.015	<0.001	0.014	<0.001	0.018	<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 NCA 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 Thurston County. Predictive validity measures how well the PSA classifies individuals in Thurston such that we would expect individuals receiving lower PSA scores to have higher rates of success (i.e., making their court date, not being rearrested, and not being rearrested for a violent crime). 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 the Thurston County data.

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 Thurston County, suggesting that for Thurston County data the PSA assessment scores do not exhibit predictive bias with respect to race and sex.

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