Bernalillo County Public Safety Assessment Validation Study

June 2021

Prepared by:
Elise Ferguson, M.A.
Helen De La Cerda, B.A.
Paul Guerin, Ph.D.
Cristopher Moore, Ph.D.

Prepared for:
New Mexico Administrative Office
of the Courts and Bernalillo County

Table of Contents

Introduction

Bernalillo County implemented the Public Safety Assessment (PSA) in June 2017. The PSA was developed by Arnold Ventures in partnership with leading criminal justice researchers (Arnold Ventures, 2019). The PSA is designed to help improve pre-trial release decision making and reduce burdens on jails.

Approximately 750,000 cases from about 300 jurisdictions across the United States were used to create the PSA (AdvancingPretrial.org, 2020). The tool was validated for over half a million cases nationally and has been re-validated at locations such as Kentucky (DeMichele et al., 2018), McLean County, Illinois (Greiner et al., 2020b); Harris County, TX (Greiner et al., 2020a); and Kane County, Illinois (Greiner et al., 2021).

The pretrial phase of a criminal case takes place from the beginning of the court case, after arrest, through the final disposition of the case, which may include a finding of guilt or innocence or a dismissal among other outcomes. During the pretrial phase, the individual is presumed innocent and is entitled to due process of law (U.S. Constitution, amend. V) and is also entitled to reasonable conditions of release, specifically either "bailable by sufficient sureties" or via relief through the courts for those unable to post a bond (N.M. Constitution, art. II, §13). Historically, release decisions have been based on the seriousness of the crime and prior criminal history and the initial conditions of release on a case were based on a jailhouse bond related to the case charges and then updated at an initial appearance. The jailhouse bonds were discontinued in 2017 shortly after the implementation of preventive detention.

Judges are often required to make decisions such as pretrial release decisions in a short period of time with incomplete information (Steffensmeier et al., 1998). Limited information coupled with limited time for decision-making, can result in disparate treatment of minorities and the poor during their pretrial period, release decision-making, and pretrial incarceration (Schlesinger, 2005). As limitations and challenges in pretrial release decision-making have become more prominent, there has been an increased interest in the development and use of pretrial risk assessment instruments (Pretrial Justice Institute, 2015).

The PSA is a judicial decision-making tool to help judges gauge the risk an individual poses and informs, rather than replaces, judicial discretion. The PSA uses evidence-based, neutral information to predict the likelihood that an individual will commit a new crime if released before trial (New Criminal Activity [NCA]), and to predict the likelihood that he/she will fail to return for a future court hearing (Failure to Appear [FTA]). In addition, it flags those individuals who present an elevated risk of committing a violent crime (New Violent Criminal Activity [NVCA]). The FTA, NCA, and NVCA are referred to as the outcome measures. By extension, these outcome measures are related to measures of pretrial success which are the inverse of each outcome measure. The inverse of the FTA Rate is the Appearance Rate and the inverse of the NCA Rate is the Public Safety Rate. As the focus of the PSA turns more towards pretrial success rather than pretrial failure, there will be continued shifts in how information is presented and structured. In this report while we focus on reporting the FTA, NCA, and NVCA because this is how the measures are calculated we realize the value and recognize the importance of focusing on how pretrial individuals succeed. When possible, we include the success rates throughout this report. As with other studies, individuals make court appearances and remain arrest free at a high rate. With this understanding, it is important to not only identify the rate in which individuals fail, but the rate individuals succeed.

In other jurisdictions that have implemented the PSA, PSAs are completed for individuals facing misdemeanor charges as well as individuals facing felony charges. In Bernalillo County, PSAs are completed for individuals with felony charges only. Felony charges are more serious than misdemeanor charges and individuals may have higher FTA and NCA scores, longer periods of pretrial release, and higher failure rates in Bernalillo County. Furthermore, each jurisdiction and Implementation Team establish what is considered violent charges and modify the Decision-Making Framework (DMF). For this reason, the findings in Bernalillo County are not comparable to other jurisdictions that have implemented the PSA.

The primary goal of this report is to validate the PSA in Bernalillo County, which would assess how accurately the PSA identifies individuals with greater probabilities of failing to appear in court and or committing new crimes, as well as across racial/ethnic groups and by gender. We also review the PSA outcome measures of FTA, NCA and NVCA, and the Adherence Rate - the degree to which conditions of release (COR) ordered corresponded with the PSA recommendation category. The Adherence Rate is compared to the outcome measures to determine if using COR in line with the PSA recommendation reduces FTAs and NCAs and their inverse measures, the Appearance Rate and Public Safety Rate. In addition, the Release Rate was calculated for the outcome measures sample and these measures are defined as follows:

- Appearance Rate – the rate at which individuals attend all scheduled court appearances[1]
- Public Safety Rate – the rate at which individuals do not have a new offense during the pretrial stage of a case
- Release Rate – the rate at which individuals are released during the pretrial stage of a case and have exposure to fail or succeed

Background on PSA Studies in Bernalillo County

Prior to this validation study, we completed several other reports that set the stage for this study. This has included a study of the Adherence Rate for the Bernalillo County Metropolitan Court (BCMC) and Second Judicial District Court (SJDC) for June 2017 through November 2017 (Institute for Social Research, 2018). We found the Adherence Rate for the BCMC was 65.6% and for the SJDC, 64.5%. Generally, in the BCMC, the COR are set at the felony first appearance (FFA) for cases with a PSA recommendation of either ROR or ROR with PTS, corresponded with their recommendation. As the level of supervision increased, the judicial decision was less likely to adhere to the PSA recommendation. In the SJDC, the COR are set at the felony arraignment (FA) for cases with a PSA recommendation of either ROR or ROR with PTS, either corresponded with the recommendation or included more restrictive conditions than recommended by the PSA. As the PTS recommended level increased, the percent of cases for which conditions were added beyond the recommendation level of the PSA also increased.

In June 2018, we completed a preliminary review of the PSA (Ferguson et al., 2018) that reported pretrial outcome results for court cases with an assessment in the BCMC and/or the SJDC between June 12, 2017 and December 31, 2017. As noted earlier, while the PSA is designed to be used for release decision-making for all arrests resulting in a booking in a jail, Bernalillo County does not use the PSA for release decision-making for misdemeanor cases, but only for felony cases. Importantly, this study reported on the use of the PSA at the point of implementation rather than waiting for a period of time

---

[1] In general, a warrant is typically issued if a defendant does not attend a hearing, but in some instances, this may not occur. For this study, the issuance of a warrant is used as the primary measure of attendance of the hearing.

to let implementation of the PSA occur; used a short time frame of 6-months; relied on a sample of both closed cases and cases pending disposition; and did not include a comparison group.

The PSA utilizes nine factors to predict failure by three pretrial outcome measures: New Criminal Activity, New Violent Criminal Activity, and Failure to Appear. For the aforementioned report, an FTA warrant was considered valid if it was issued during the study period for failure to appear at a court hearing and included an arrest or related court event that indicated the warrant was valid. The FTA for the BCMC cases was .4%. BCMC-SJDC cases had an FTA Rate of 20.3% and the rate for SJDC cases was 31.3%. The overall FTA Rate for all case categories was 7.0% with an Appearance Rate of 93.0%. In general, and not unexpectedly, the NCA Rate increased as the NCA score increased for different case types. The NCA Rate for the BCMC cases was 16.1%. BCMC-SJDC cases had an NCA Rate of 29.6% and the rate for SJDC cases was 21.7%. The overall NCA Rate for all case categories was 19.3% and the Public Safety Rate was 80.7%. The NVCA Rate was generally higher for cases with higher NCA scores. The NVCA Rate for the BCMC cases was 3.8%. BCMC-SJDC cases had an NVCA Rate of 8.6% and the rate for SJDC cases was 2.0%. The overall NVCA Rate for all case categories was 4.7% with a rate of 95.3% not having a new violence offense.

Of importance, the time frame for this study was short. Evaluating the PSA after being used for six months was not optimal. This short time frame did not allow for the use of the PSA to become stabilized and does not allow a long enough time period to measure the implementation or outcomes. The current review includes cases that were still open at the end of the study period. We believe a better study design would rely on closed cases and a longer period of time to ensure the proper implementation of the PSA and a large enough sample of closed cases. We recommended this study be replicated after the PSA has been in use for a minimum of one year.

In November 2019, (Ferguson et al., 2019) we completed an update to the June 2018 report that used 12 months of closed cases and reviewed the Adherence Rate and how the PSA performs on the pretrial outcomes measured by the PSA - FTA, NCA, and NVCA. In addition, this report described the charge levels for the NCAs that occurred during the pretrial period. The outcome measures for the PSA were calculated for the 6,392 cases in the study by recommendation category. As the FTA and NCA/NVCA scores increased, so did the percent of failures. The overall FTA Rate was 18.5% and an Appearance Rate of 81.5%. The FTA Rate generally increased as the FTA score from the PSA increased. The NCA Rate also increased as the NCA score increased, with an overall NCA Rate of 17.2%, and a Public Safety Rate of 82.8%. The NVCA Rate was highest among those cases with higher NCA scores, with an overall NVCA Rate of 4.1%, with 95.9% of individuals not having a new violence offense.

Cases within the sample were also reviewed for adherence. There were 697 Preventive Detention cases, where an individual was detained while awaiting a SJDC detention hearing that were excluded, and 5,695 cases were reviewed. Adherence was determined by comparing the PSA recommendation for a case to the ordered COR. A brief analysis of the Adherence Rates between July 1, 2017 and December 31, 2018, found the overall Adherence Rate in 2017 was 72.1% and 77.7% in 2018, a small increase from the earlier study on Adherence Rates.

In January 2020, we completed a study of the impact of preventive detention motions on the FTA and NCA Rate for individuals in Bernalillo County charged with felony crimes and for which the PSA was administered and used in the pretrial release decision-making process from the Bernalillo County Metropolitan Detention Center (MDC) (Ferguson et al., 2020). This review presented the FTA Rate and

NCA Rate for cases assessed using the PSA in BCMC. These outcome measures were calculated by the filing of a PTD motion on the case, the recommendation category, and the crime type.

The relationship between the FTA and NCA Rates and the PSA recommendation categories was statistically significant, indicating the knowledge of the PSA category improved the ability to estimate the likelihood of failures of either type. Similarly, the crime type was also significantly associated with the FTA and NCA Rates, with the highest rates of failure among drug offenses, property offenses, and public order/other offenses. This could be for a variety of reasons, such as a difference in the number of hearings, the length of cases, and the likelihood of indictment.

The FTA Rate for cases with a PTD motion was 17.8% and the Appearance Rate was 82.2% and for those without a PTD motion the FTA Rate was 18.1% and the Appearance Rate was 81.9%. The NCA Rate for those with a PTD motion was 18.0% with a Public Safety Rate of 82.0% and the NCA Rate for those with no motion was 16.7% and the Public Safety Rate was 83.3%. The difference between these groups was not statistically significant, indicating that the PTD motion was not associated with failure. While the PSA recommendation and crime type improved the ability to estimate the likelihood of failure, the PTD motion – or lack thereof – is not a good indicator of future failure. Instead, the PSA recommendation and crime type provided the most accurate means to estimate future failure.

For the NCAs that occurred, the new criminal activity was of a different type more than half the time for violent, drug, and property offenses, although the largest single category corresponded with the original charge. Additionally, the majority of new criminal activity was for 4th degree felonies (539 or 51.7%) followed by misdemeanors (243 or 23.3%). For the NCA, 42.2% of the cases had a lower charge than the assessed case and an additional 45.3% are of the same level. Overall, NCA occurred for fewer than one in five cases and was primarily for charges of a lower or equivalent level as the assessed case.

Literature Review

The PSA was developed to use evidence-based, neutral information and assists judges with release decision making. A number of studies of the Arnold Ventures PSA have been conducted including validation studies to determine if the PSA measures what it is intended to and does so in an equitable manner. This brief review focuses on validation studies and includes general review of the PSA. In 2018, Arnold Ventures published a guide for users of the PSA about how to define and measure outcomes, the importance of tracking PSA performance, other important measures (i.e., Adherence Rate and pretrial length of stay), the importance of conducting a validation study, and how oversight includes tracking and measuring the outcome factors.

Unlike in New Mexico, the PSA is utilized statewide in Kentucky. The Center for Safety, Justice and Resilience conducted a validation study of the PSA in Kentucky. The predictive validity, differential validity, predictive bias, and accuracy of the PSA, by race and gender were assessed for 164,597 cases. The FTA Rate, NCA Rate, and NVCA Rate were 14.8%, 10.6%, and 1.1%, respectively. One finding included evidence that as the PSA scores increased, so did failure rates. The Area Under the Curve Receiver Operating Characteristics (AUC ROC) probability curve values were highest for the NVCA flag (AUC = 0.664), followed by the NCA score (AUC = 0.650), and finally the FTA scale (AUC = 0.646). This demonstrated predictive validity. Next, the FTA scores demonstrated more predictability for White individuals (AUC = 0.655) than Black individuals (AUC = 0.612). The NCA was slightly more predictive for Black individuals (AUC = 0.659) than White individuals (AUC = 0.647). The FTA and NCA scores differed

by gender as well - the FTA scale was more predictive for females (AUC = 0.655 and 0.642 for males), and the NCA scale was more predictive for males. (AUC = 0.653 and 0.637 for females). Overall, White individuals (AUC = 0.666) were 14% more likely to have NCA and Black individuals had slightly higher NVCA Rates (AUC = 0.631. There was no predictive bias for FTA or NCA Rates across gender (DeMichele et al., 2018).

Recently, the Access to Justice Lab validated the PSA in McLean County, Illinois, Kane County, Illinois, and Harris County, Texas. In McLean County, data was provided by the Integrated Justice Information System (IJIS) which consists of data from McLean County Pretrial Unit, jail release data from Sheriff's Department, and case information from the court. There were 6,345 cases for some 4,707 unique individuals. The FTA Rate was 18.1%, the NCA Rate was 15.5%, and the NVCA Rate was 5.5%. As the FTA and NCA scores increased, so did failure rates. The NCA score was uniformly valid but the FTA score was not. The equitable validity was assessed, or whether the PSA validates equally across race and gender. It was demonstrated that males commit more NCA and NVCA at higher rates than females, and FTA at lower rates than females. Next, White individuals failed at lower rates than Black individuals across all three outcomes (Greiner et al., 2020b).

In Harris County, Texas, data was provided to the Access to Justice Lab by way of query and data pulls from multiple agencies: Harris County District Clerk, Harris County, Justice Administration Department, and Harris County Pretrial. Ultimately, the sample consisted of 61,303 cases. The overall validity, uniform validity, and equitable validity were assessed. The FTA Rate was 18.8%, NCA Rate was 17.4%, and the NVCA Rate was 4.1%. The AUC values were highest for NCA (0.66), then 0.60 for FTA, and 0.55 for NVCA. Bivariate regression tests supported equitable validity for FTA and NCA, but no evidence of uniform validity. The AUC values were highest for FTA, at 0.70, then NCA at 0.67, and NVCA at 0.61. It was found that the rates at which Black individuals commit NCA were much higher than White individuals. Failure rates were lower for female individuals than males across all three outcomes (Greiner et al., 2020a).

In Kane County, Illinois, data was provided to the Access to Justice Lab by Kane County Circuit Clerk, Kane County Sheriff's Department, and Kane County Court Services. There were 13,094 assessments for 10,059 unique individuals that were either charged with misdemeanors and felonies. As with McLean County and Harris County, the overall validity, uniform validity, and equitable validity, were assessed. The FTA Rate was 22.3%, the NCA Rate was 16.6%, and the NVCA Rate was 7.1%. The AUC values were highest for FTA at 0.67, followed by NCA at 0.63, and 0.61 for NVCA. By way of bivariate logistic regression modeling, there was strong evidence for the overall validity of the PSA. Regarding uniform validity, logistic regression provided evidence for FTA outcomes only. For NCA, NVCA, and FTA, there were no significant differences between racial groups and there was weak evidence that the PSA does not equitably validate. Overall, Black individuals had higher rates for all three outcomes and males had higher NCA and NVCA Rates than females (Greiner et al., 2021).

There are limited validation studies of the PSA and a prominent issue is validating the PSA for each jurisdiction in which it has been implemented. Each jurisdiction decides who will be involved in the implementation, as well as modifying COR or recommendations that are used in the DMF. Additionally, it is important to look at local pretrial, jail, and court processes. Validation studies are vital, and more work should be done on analyzing the predictability of the tool. However, each validation study and the findings are unique to the jurisdiction. Therefore, the validation of the PSA in Bernalillo County may differ from findings in other jurisdictions.

PSA Implementation in Bernalillo County

Each jurisdiction that utilizes the PSA develops a DMF. The DMF plots the individual's NCA score against the individual's FTA score and provides a recommended release category for the individual. In early 2017, criminal justice stakeholders in Bernalillo County formed a PSA Implementation Team. The Team included representatives from criminal justice entities that would be impacted by the PSA: the BCMC, the SJDC, the Office of the Public Defender, Offices of the District Attorney and Public Defender, Pretrial Services, and local law enforcement agencies. The PSA Implementation Team met regularly to prepare for the use of the PSA in Bernalillo County. The Team developed the Decision-Making Framework for Bernalillo County.

The PSA is scored by reviewing an individual's criminal history, current cases, and age to create an FTA score and an NCA score as well as a flag for NVCA. Table 1 below combines the relationship between the risk factors and the three pretrial outcome measures (LJAF, 2016). An "X" indicates an increase in the individual's likelihood of that outcome, based on the risk factor. For instance, if the individual's current offense is violent (#2 below), it increases that individual's likelihood of committing an NVCA during their pretrial period.

Table 1. PSA Risk Factors and Pretrial Outcomes			
Risk Factor	**Pretrial Outcome**		
	FTA	**NCA**	**NVCA**
1. Age at current arrest		X	
2. Current violent offense			X
2A. Current violent offense and 20 years old or younger			X
3. Pending charge at the time of the offense	X	X	X
4. Prior misdemeanor conviction		X	
5. Prior felony conviction		X	
5A. Prior conviction (misdemeanor or felony)	X		X
6. Prior violent conviction		X	X
7. Prior failure to appear in the past two years	X	X	
8. Prior failure to appear older than two years	X		
9. Prior sentence to incarceration		X	

Using the FTA and NCA scales, a release recommendation for each case is assigned using the DMF. The DMF provides recommendations that range from release on own recognizance (ROR), various levels of pre-trial supervision, up to a recommendation to detain or release with maximum conditions. The level of pre-trial supervision, or pre-trial monitoring level (PML) ranges from level 1 to level 4 with increasing degrees of supervision and conditions as the level increases[2]. The DMF was tailored for this jurisdiction by a PSA Implementation Team involving key stakeholders (see Table 2).

---

[2] The PML level indicates to what degree the defendant should be supervised, including the frequency and type of contacts with PTS (phone versus office visits), electronic monitoring, and the frequency of UAs.

Table 2. Decision-Making Framework in Bernalillo County						
		New Criminal Activity Scale				
		NCA 1	NCA 2	NCA 3	NCA 4	NCA 5	NCA 6
Failure to Appear Scale	FTA 1	(A) ROR	(B) ROR				
	FTA 2	(C) ROR	(D) ROR	(E) ROR PML 1	(F) ROR PML 3	(G) ROR PML 4	
	FTA 3		(H) ROR PML 1	(I) ROR PML 2	(J) ROR PML 3	(K) ROR PML 4	(L) Detain or Max Conditions
	FTA 4		(M) ROR PML 1	(N) ROR PML 2	(O) ROR PML 3	(P) ROR PML 4	(Q) Detain or Max Conditions
	FTA 5		(R) ROR PML 2	(S) ROR PML 2	(T) ROR PML 3	(U) Detain or Max Conditions	(V) Detain or Max Conditions
	FTA 6				(W) Detain or Max Conditions	(X) Detain or Max Conditions	(Y) Detain or Max Conditions

The next several sections report the sample, the PSA recommendation categories and the PSA outcome measures: the FTA Rate, the NCA Rate, and the NVCA Rate. A failure to appear warrant was considered valid if it was issued during the study period for failure to appear at a court hearing[3]. New criminal activity does not include City or County ordinances or traffic offenses per decisions made with local stakeholders and Arnold Ventures. Stakeholders also identified violent criminal activity, which consists of offenses such as murder, kidnapping, or any offense or conspiracy to commit such offense, which causes physical injury to another person.[4]

Objectives

The purpose of this study is to assess the validity of the PSA, specifically whether the PSA measures what it is intended to, which is predicting the likelihood of an individual failing to appear at future court hearings (FTA), committing new criminal activity (NCA), and new violent criminal activity (NVCA), during their pretrial period. Overall validity, predictive validity, uniform validity, and equitable validity are assessed using Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC), Chi-Square Test of Independence, Pearson's Correlation, and Odds Ratio from Logistic Regression. For purposes of this study, overall validity is the comparison of failure rates to risk factor scores. Uniform validity is the comparison of increased risk factor scores to an increase in failure rates. Predictive validity is the comparison of PSA scores to failure rates. Finally, equitable validity looks at whether the PSA is equally valid for race and gender.

In addition to the validation portion of this study, failure rates are reported by PSA recommendation category, by FTA or NCA score, and by race and gender. For cases with an NCA, types of charges are reported by charge level and type. Finally, adherence, or how often the ordered COR (COR) match the PSA recommendation, is reported by recommendation category, by scores, and race and gender.

---

[3] While in some circumstances the warrant is issued in error or a cancellation is issued based on stipulation with the courts, this study does not distinguish between the two.

[4] The New Mexico Criminal Code was reviewed and violent offenses categorized by committee prior to the implementation of the PSA. It is worth noting that some statutes may have subsections which are not considered violent offenses.

Sample

The sample was constructed using electronic court data from the Bernalillo County Metropolitan Court and the Second Judicial District Court. The electronic data included all cases filed in the time frame. From this data, cases for the sample were selected if:

1. The case was opened between July 1, 2017 and March 31, 2020;
2. There was a PSA completed;
3. The individual was in custody for either the Felony First Appearance (FFA) or the Felony Arraignment (FA);
4. The case was closed and no longer pending on March 31, 2020; and
5. The individual was released during the pretrial period between FFA or FA and final case disposition.

Between July 1, 2017 and March 31, 2020, 29,645 cases were opened, 18,285 in BCMC and 11,360 in SJDC. For a variety of reasons, 11,644 (39.3%) were excluded from analysis. Fugitive cases, or cases where the individual is facing charges from another state or jurisdiction (722 cases), were excluded as they are not considered eligible for assessment (see Table 3). Cases in the sample were categorized by case filings and the indictments. Oftentimes, there was overlap between the BCMC case and a SJDC case filed during the pretrial period. In Bernalillo County, most felony cases begin in BCMC with few exceptions. Once a case was filed, the prosecution had 60 days from the individual's FFA date and pretrial release date to charge the individual through either a grand jury indictment or through preliminary hearing where a judge may decide there exists evidence sufficient to indict. Once indicted, the BCMC case is linked to the SJDC case and the case continued in SJDC. For the purposes of the study, the overlapping BCMC and SJDC case were considered one case. If the indictment occurred after the BCMC pretrial period, the SJDC case is considered separately, with its own pretrial period. The findings from this study are reported in the aggregate rather than separately by court. There were 4,470 cases excluded because the SJDC portion of the case was attached to the BCMC portion.

An additional 5,199 cases were excluded because they had no PSA. This occurs most often for SJDC cases, as when a case is indicted that does not necessarily lead to a booking prior to a felony arraignment. Of the remaining cases, 1,185 were excluded because the inmate was not in custody for the felony first appearance or had an FTA warrant prior to the felony arraignment. The remaining 68 cases had a variety of issues including no felony arraignment or no COR.

Table 3. Exclusions from Electronic Court Data			
Exclusion Reason	**BCMC**	**SJDC**	**Total**
Fugitive	722	0	722
In the BCMC-SJDC File	0	4,470	4,470
No PSA / not in custody for PSA	440	4,759	5,199
Not in custody for felony first appearance / FTA at or before felony arraignment	318	867	1,185
Other (No felony arraignment, no conditions of release, etc.)	0	68	68
Total	1,480	10,164	11,644

The remaining 18,001 cases met the criteria of having a PSA and the individual was in custody for the release decision.

Case Status

For each case in the sample, the case status was collected and identified as either closed or pending. A case was considered closed if there was a closing event or final disposition, such as a sentence, dismissal, plea bargain, or finding of no probable cause, on or before March 31, 2020. If the case had a closing event, then the date of the event was collected as the case close date. If the court case did not have a closing event or disposition, a 60-day proxy close date was calculated. This date is the deadline for which the individual must be indicted, or the case is dismissed. In these cases, the case proxy close date was calculated 60 days from the FFA hearing date, and this was considered the case close date. If the individual was indicted within the 60 days, it was considered a BCMC-SJDC case. If the indictment occurred outside of the 60 days or did not occur at all, the case was considered BCMC only. A case was identified as pending if there was not a closing event or disposition by March 31, 2020. Only cases that were both opened and closed between July 1, 2017, and March 31, 2020, were included in this analysis.

Of the 17,978 cases with an individual in custody for release decision making and an assessment, 2,070 (11.5% of eligible cases) were pending. The remaining 15,908 closed cases were reviewed for exposure.

Exposure

Exposure is used to indicate when an individual spends time in the community during the pretrial period of the assessed case. Jail booking and release data from the MDC were merged to the individual's court case and the corresponding booking for the assessment and hearing at which the release decision was made was selected. If an individual was not released during the pretrial portion of the case or was released to the New Mexico Corrections Department (NMCD) they were identified as having no exposure. Inmates with exposure had the opportunity to have an FTA or to commit a new crime. Of the 15,908 closed cases, 5,619 did not have exposure during the pretrial period. The remaining 10,289 had exposure during the pretrial period.

The Release Rate for the current sample was 69.1%[5]. While the Release Rate was 2.6% lower than the Release Rate in the report for cases through March 2018 (Ferguson et al., 2019), these figures are not directly comparable due to a variety of changes in methodology between the two studies. There were some minor changes to how case closing events were classified, resulting in some cases which had previously been identified as having exposure now being listed as having no exposure. Cases which were closed in lieu of COR at the felony first appearance or felony arraignment were also removed from analysis which adjusted the Release Rate. In addition, of the assessed cases for which the individual was in custody for the release decision, the portion of cases which were pending was much higher for the prior sample (20.2%) compared to the current sample (11.5%).

The 10,289 assessed, closed cases from BCMC and SJDC with exposure comprised the sample of cases for reviewing outcome measures.

---

[5] The Release Rate varied greatly by collapsed PSA recommendation categories. Cases with an ROR recommendation had a Release Rate of 93.6%, ROR – PML 1 85.6%, ROR – PML 2 76.9%, ROR – PML 3 65.0%, ROR – PML 4 59.5%, and Detain/Max 43.8%.

Sample Description

Basic demographics, including gender, race, and ethnicity, were collected from MDC booking data. The majority of cases had a male defendant (74.5%) (see Table 4). In just over half of the cases, the defendant was Hispanic (50.4%). The percentage of White defendants was 32.4%, followed by Native American and Black defendants at 7.1% and 6.9% respectively. The remaining 3.1% of defendants for cases in the sample were unknown, multi-racial, or another race category.

Table 4. Gender and Race in Outcome Sample		
Category	**Count**	**Percent (N=10,289)**
Gender		
Male	7,670	74.5%
Female	2,619	25.5%
Race		
Hispanic	5,188	50.4%
White	3,334	32.4%
Native American	733	7.1%
Black	713	6.9%
Other	321	3.1%

The NCA and FTA scores result in a PSA score on the DMF, of which there can be 25 options for release recommendations. The recommendations from the DMF can be either of the following collapsed PSA recommendation categories: ROR, ROR – PML 1, ROR – PML 2, ROR – PML 3, ROR – PML 4, or detain or release with maximum conditions (hereafter Detain/Max).

Table 5 shows the recommendation categories. The most commonly assigned categories included 12.3% or 1,264 in the (A) ROR category, 10.3% or 1,057 in the (I) ROR – PML 2 category, and 8.6% or 884 in the (J) ROR – PML 3 category. The categories that were least assigned included 2 (0.02%) in the (G) ROR PML 4 category, 5 (0.05%) in the (L) Detain/Max category, and 20 (0.2) in (R) ROR – PML 2 category.

Table 5. PSA Recommendation Categories													
		New Criminal Activity Scale											
		NCA 1		**NCA 2**		**NCA 3**		**NCA 4**		**NCA 5**		**NCA 6**	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
FTA Scale	**FTA 1**	(A) ROR		(B) ROR									
		1,264	12.3%	713	6.9%								
	FTA 2	(C) ROR		(D) ROR		(E) ROR PML 1		(F) ROR PML 3		(G) ROR PML 4			
		225	2.2%	638	6.2%	606	5.9%	332	3.2%	2	0.02%		
	FTA 3			(H) ROR PML 1		(I) ROR PML 2		(J) ROR PML 3		(K) ROR PML 4		(L) Detain or Max Conditions	
				618	6.0%	1,057	10.3%	884	8.6%	84	0.8%	5	0.05%
	FTA 4			(M) ROR PML 1		(N) ROR PML 2		(O) ROR PML 3		(P) ROR PML 4		(Q) Detain or Max Conditions	
				193	1.9%	444	4.3%	525	5.1%	399	3.9%	74	0.7%
	FTA 5			(R) ROR PML 2		(S) ROR PML 2		(T) ROR PML 3		(U) Detain or Max Conditions		(V) Detain or Max Conditions	
				20	0.2%	184	1.8%	533	5.2%	508	4.9%	257	2.5%
	FTA 6							(W) Detain or Max Conditions		(X) Detain or Max Conditions		(Y) Detain or Max Conditions	
								144	1.4%	156	1.5%	424	4.1%

Table 6 shows the collapsed PSA recommendation categories for the sample. The most frequently recommended category was ROR, which accounted for 2,840 cases (27.6%). Additionally, there were 2,274 or 22.1% that had ROR – PML 3 recommended. The Detain/Max and ROR – PML 2 categories consisted of 15.2% and 16.6%, or 1,568 and 1,705 cases, respectively. The two smallest collapsed categories were ROR – PML 1 (1,417 cases or 13.8%) and ROR – PML 4, with 4.7% of cases (485).

Table 6. PSA Collapsed Recommendation Category		
Release Category	Count	Percent
ROR	2,840	27.6%
ROR - PML 1	1,417	13.8%
ROR - PML 2	1,705	16.6%
ROR - PML 3	2,274	22.1%
ROR - PML 4	485	4.7%
Detain/Max	1,568	15.2%
Total	10,289	100.0%

The next section reports the FTA, NCA, and NVCA outcome measures for the 10,289 cases in the sample.

PSA Risk Factors and Scores

This section reports on the three outcomes measured by the PSA - Failure to Appear (FTA), New Criminal Activity (NCA), and New Violent Criminal Activity (NVCA). In addition, this section includes details on the charge level for the NCA that occurred during the pretrial period.

Failure to Appear and Risk Factors

Table 7 displays the risk factors that determine the individual's FTA score. There were 29.8% or 3,064 individuals who had a pending charge at the time of arrest. In just over two thirds of cases, the individual had a prior misdemeanor or felony conviction (69.0%). The number of cases where the individual had no prior FTAs in the past two years was 6,782 or 65.9%. There were 1,702 (16.5%) cases where the individual had one FTA and 1,805 (17.5%) where the individual had two or more FTAs in the past two years. In just over half of the cases, the individual had an FTA older than 2 years (51.6%).

Table 7. FTA Risk Factors			
Factor	Response	Count	Percent (N=10,289)
Pending Charge at Time of Arrest	No	7,225	70.2%
	Yes	3,064	29.8%
Any Prior Conviction (MD or Felony)	No	3,187	31.0%
	Yes	7,102	69.0%
Prior FTA in Past 2 Years	0	6,782	65.9%
	1	1,702	16.5%
	2 or More	1,805	17.5%
Prior FTA Older Than 2 Years	No	4,983	48.4%
	Yes	5,306	51.6%

Table 8 displays the Pearson's correlation coefficients (r) for the four factors used to calculate the FTA and the overall FTA score. The correlations for all four of the factors are positive and small. These correlations were all statistically significant (p<.001[6]). An FTA during the prior 2 years had the strongest correlation (r=.176) and having other pending charges at the time of arrest had the lowest correlation (r=0.122). The overall correlation between the FTA score and the FTA outcome was r=.212.

Table 8. Pearson Correlation Coefficients for FTA Risk Factors and Outcomes		
Factor	Pearson's r	Standard Error
Pending Charge at Time of Arrest	0.122*	+/- 0.010
Any Prior Conviction (MD or Felony)	0.124*	+/- 0.009
Prior FTA in Past 2 Years	0.176*	+/- 0.011
Prior FTA Older Than 2 Years	0.131*	+/- 0.010
FTA Score	0.212*	+/- 0.010

*p<.001

Table 9 displays the frequency of cases by the calculated FTA score. The largest number and percent of cases received an FTA 3 score (25.7%). FTA 1 and FTA 2 accounted for 19.2% and 17.5% of the sample, respectively. FTA 4 and FTA 5 accounted for 15.9% and 14.6% of the sample. The score with the smallest percent of cases was FTA 6, with 724 individuals (7.0%).

Table 9. FTA Score for Outcome Measures Sample		
FTA Score	Count	Percent
FTA 1	1,977	19.2%
FTA 2	1,803	17.5%
FTA 3	2,648	25.7%
FTA 4	1,635	15.9%
FTA 5	1,502	14.6%
FTA 6	724	7.0%
Total	10,289	100.0%

New Criminal Activity and Risk Factors

Table 10 displays the frequency and percent responses to the risk factors used to calculate the NCA score. The large majority of cases had a defendant who was 23 or older (87.1%) and the majority did not have other pending charges at the time of arrest (70.2%). In nearly two thirds (64.2%) of cases the individual had a prior misdemeanor conviction and in 38.1% of cases the individual had a prior felony conviction. In 70% of the cases, the individual had no prior violent convictions, in 23.4% of cases the individual had 1 or 2 violent convictions and in the remaining 6.6% of cases the individual had 3 or more violent convictions. In slightly more than half of the cases, the individual did not have a prior sentence that included incarceration (55.7%).

---

[6] Similar to the Kentucky study (DeMichele et al.), we conservatively set the threshold p-value for statistical significance at 0.001. Some readers may be accustomed to a higher threshold of 0.05, but since we perform more than 60 tests of accuracy and fairness on the PSA across different outcomes, demographic groups, and score categories, using 0.05 would cause us to incorrectly claim that some differences are significant when they are actually just noise in the data (technically, it would cause Type I errors where we incorrectly reject a null hypothesis that no difference exists). See for instance https://en.wikipedia.org/wiki/Bonferroni_correction

Table 10. NCA Risk Factors			
Factor	**Response**	**Count**	**Percent (N=10,289)**
Age at Current Arrest	22 or Younger	1,323	12.9%
	23 or older	8,966	87.1%
Pending Charge at Time of Arrest	No	7,225	70.2%
	Yes	3,064	29.8%
Prior Misdemeanor Conviction	No	3,679	35.7%
	Yes	6,610	64.2%
Prior Felony Conviction	No	6,372	61.9%
	Yes	3,917	38.1%
Prior Violent Conviction	0 Violent Convictions	7,199	70.0%
	1 or 2 Violent Convictions	2,408	23.4%
	3 or More Violent Convictions	682	6.6%
Prior FTA in Past 2 Years	0	6,782	65.9%
	1	1,702	16.5%
	2 or More	1,805	17.5%
Prior Sentence to Incarceration	No	5,728	55.7%
	Yes	4,561	44.3%

Table 11 displays the Pearson's correlation coefficients (r) for the seven factors used to calculate the NCA and the overall NCA score. The correlations for all seven of the factors are positive and small. These correlations were all statistically significant (p<.001) with the exception of age at current arrest. An FTA during the prior 2 years had the strongest correlation (r=.155) followed by pending charges at the time of arrest (r=.130). The overall correlation between the NCA score and the NCA outcome was r=.191.

Table 11. Pearson Correlation Coefficients for NCA Risk Factors and Outcomes		
Factor	**Pearson's r**	**Standard Error**
Age at Current Arrest	0.013	+/- 0.010
Pending Charge at Time of Arrest	0.130*	+/- 0.010
Prior Misdemeanor Conviction	0.082*	+/- 0.009
Prior Felony Conviction	0.095*	+/- 0.010
Prior Violent Conviction	0.067*	+/- 0.010
Prior FTA in Past 2 Years	0.155*	+/- 0.011
Prior Sentence to Incarceration	0.112*	+/- 0.010
NCA Score	0.191*	+/- 0.009

*p<.001

NCA scores were more evenly distributed than the FTA scores, with NCA 2 (21.2%), NCA 3 (22.3%), and NCA 4 (23.5%), comprising the majority of cases (see Table 12). NCA 1 accounted for 14.5% of cases and NCA 5 accounted for 11.2% of cases. NCA 6 had the smallest number of cases, 760 or 7.4%.

Table 12. NCA Score for Outcome Measures Sample		
NCA Score	**Count**	**Percent**
NCA 1	1,489	14.5%
NCA 2	2,182	21.2%
NCA 3	2,291	22.3%
NCA 4	2,418	23.5%
NCA 5	1,149	11.2%
NCA 6	760	7.4%
Total	10,291	100.0%

New Violent Criminal Activity and Risk Factors

Table 13 displays the risk factors that determine whether the individual receives the NVCA flag. In just over a third of cases (38.3%) the individual's current offense was violent and in 2.8% the individual was also 20 years old or younger. In 29.8% of cases the individual had another charge pending and in 69.0% of cases the individual had a prior misdemeanor or felony conviction. In most cases (70.0%) the individual had no prior violent convictions, in 23.4% of cases the individual had 1 or 2 violent convictions and in the remaining 6.6% of cases the individual had 3 or more violent convictions.

Table 13. NVCA Risk Factors			
Factor	**Response**	**Count**	**Percent (N=10,289)**
Current Violent Offense	No	6,350	61.7%
	Yes	3,939	38.3%
Current Violent Offense and ≤ 20 Years Old	No	9,998	97.2%
	Yes	291	2.8%
Pending Charge	No	7,225	70.2%
	Yes	3,064	29.8%
Any Prior Conviction (MD or Felony)	No	3,187	31.0%
	Yes	7,102	69.0%
Prior Violent Conviction	0 Violent Convictions	7,199	70.0%
	1 or 2 Violent Convictions	2,408	23.4%
	3 or More Violent Convictions	682	6.6%

Table 14 displays the Pearson's correlation coefficients (r) for the five factors used to calculate the NVCA and the overall NVCA score. The correlations for all five of the factors are positive and small. These correlations were statistically significant (p<.001). The strongest correlations were for current violent offense factor and (r=.068) prior violent convictions (r=.063). The overall correlation between the NVCA score[7] and the NVCA outcome was r=.083.

Table 14. Pearson Correlation Coefficients for NVCA Risk Factors and Outcomes		
Factor	**Pearson's r**	**Standard Error**
Current Violent Offense	0.068*	+/- 0.010
Current Violent Offense and ≤ 20 Years Old	0.026	+/- 0.013
Pending Charge	0.052*	+/- 0.011
Any Prior Conviction (MD or Felony)	0.029	+/- 0.009
Prior Violent Conviction	0.063*	+/- 0.011
NVCA Flag	0.083*	+/- 0.012

*p<.001

Of the 10,289 individuals in the sample, 16.0% (1,650 cases) had a new violent criminal activity flag (see Table 15). The remaining 8,639 (84.0%) did not have the violence flag.

Table 15. NVCA Flag for Outcome Measures Sample		
Flag	**Count**	**Percent**
No NVCA Flag	8,639	84.0%
NVCA Flag	1,650	16.0%
Total	10,289	100.0%

---

[7] The raw NVCA score had a higher Pearson's correlation value (r=.108, p<.001) than the NVCA flag, but this score is not provided for release decision-making.

PSA Outcome Measures

The three outcomes measured by the PSA include Failure to Appear (FTA), New Criminal Activity (NCA), and New Violent Criminal Activity (NVCA). This section will consist of details of each outcome measures, as well as charge levels for the NCA that was committed within the pretrial period.

Failure to Appear Rates

Any warrant issued for FTA at a scheduled court appearance was considered a valid warrant and applied to the FTA Rate[8]. The overall FTA Rate was 22.9% and an Appearance Rate of 77.1%. Table 16 below displays the FTA Rate by the PSA recommendation category. The highest FTA Rates were observed in the higher recommendation categories and the lowest rates were observed in the lower categories.[9] Specifically, the rates were highest in the (X) Detain/Max and (Y) Detain/Max categories, with 71 out of 156 (45.5%), and 179 out of 424 (42.2%). The next highest FTA Rates were found in category (W) Detain/Max (39.6%) and (V) Detain/Max (37.7%). The lowest FTA Rates were seen in the (A) ROR and (D) ROR categories, with 135 out of 1,264 (10.7%) and 64 out of 638 (10.0%).

Table 16. FTA Rates by PSA Recommendation Category														
		New Criminal Activity Scale												
		NCA 1		NCA 2		NCA 3		NCA 4		NCA 5		NCA 6	
		Count	FTA Rate	Count	FTA Rate	Count	FTA Rate	Count	FTA Rate	Count	FTA Rate	Count	FTA Rate
FTA Scale	FTA 1	(A) ROR 135	10.7%	(B) ROR 78	10.9%								
	FTA 2	(C) ROR 38	16.9%	(D) ROR 64	10.0%	(E) ROR PML 1 116	19.1%	(F) ROR PML 3 66	19.9%	(G) ROR PML 4[10] 1	50.0%		
	FTA 3			(H) ROR PML 1 97	15.7%	(I) ROR PML 2 259	24.5%	(J) ROR PML 3 234	26.5%	(K) ROR PML 4 19	22.6%	(L) Detain or Max Conditions 2	40.0%
	FTA 4			(M) ROR PML 1 46	23.8%	(N) ROR PML 2 108	24.3%	(O) ROR PML 3 161	30.7%	(P) ROR PML 4 121	30.3%	(Q) Detain or Max Conditions 22	29.7%
	FTA 5			(R) ROR PML 2 4	20.0%	(S) ROR PML 2 56	30.4%	(T) ROR PML 3 176	33.0%	(U) Detain or Max Conditions 148	29.1%	(V) Detain or Max Conditions 97	37.7%
	FTA 6					(W) Detain or Max Conditions 57	39.6%	(X) Detain or Max Conditions 71	45.5%	(Y) Detain or Max Conditions 179	42.2%		

Figure 1 displays the FTA Rate with associated 95% confidence intervals by FTA score. As the FTA scores increased, the FTA Rate increased and individuals who scored a 1 on the FTA scale failed to appear 10.8% of the time. The FTA Rate then increased to 15.8% for those with a 2, then 23.1% for an FTA 3, 28.0% for an FTA 4, 32.0% for an FTA 5, and finally 42.4% for those with the highest score of 6. The Appearance Rate ranged from 89.2% for cases with an FTA 1 down to 57.6% for cases FTA 6.

---

[8] While any warrant for a failure to appear issued has typically been considered a valid warrant, as definitions are refined over time, there may be future attempts to distinguish between willful and unintentional failures to appear. This is difficult to ascertain and would require manual review of court documents.

[9] Categories G and L have such a small number of cases that a representative FTA Rate cannot be reliably calculated.

Figure 1. FTA Rate by FTA Score

FTA Rate by Score · · · · · · Overall FTA Rate (22.9%)

- FTA 1: 10.8%
- FTA 2: 15.8%
- FTA 3: 23.1%
- FTA 4: 28.0%
- FTA 5: 32.0%
- FTA 6: 42.4%
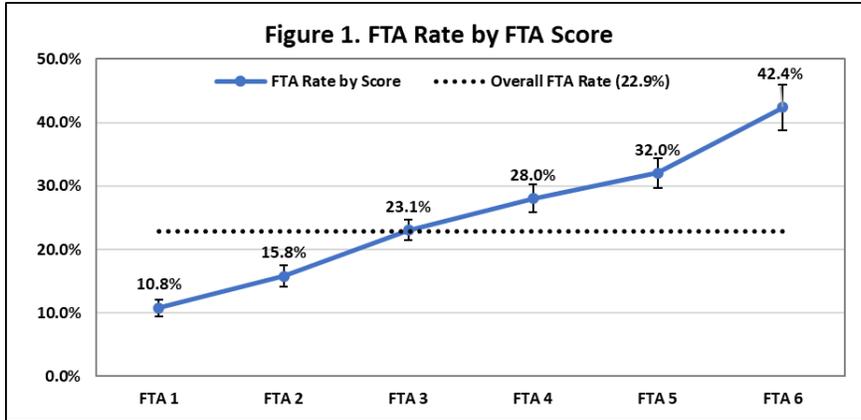
Figure 2 below displays the FTA Rate by the collapsed recommendation category. Overall, as the recommendation category became more restrictive, the FTA Rate increased generally and the Appearance Rate decreased. For individuals in the ROR, the least restrictive category, the FTA Rate was 11.1%, followed by 18.3% in ROR – PML 1, 25.0% in ROR – PML 2, 28.0% in ROR – PML 3, 29.1% for ROR – PML 4, and finally 36.7% for individuals in the Detain/Max category. There was a small increase in the FTA Rate from ROR – PML 2 to ROR – PML 3 (3.0%) and an even smaller increase from ROR – PML 3 to ROR – PML 4 (1.1%). The Appearance Rate ranged from 88.9% for cases with an ROR recommendation down to 63.3% for cases with a Detain/Max recommendation.

Figure 2. FTA Rate by Collapsed Recommendation Category

- ROR: 11.1%
- ROR - PML 1: 18.3%
- ROR - PML 2: 25.0%
- ROR - PML 3: 28.0%
- ROR - PML 4: 29.1%
- Detain Max: 36.7%
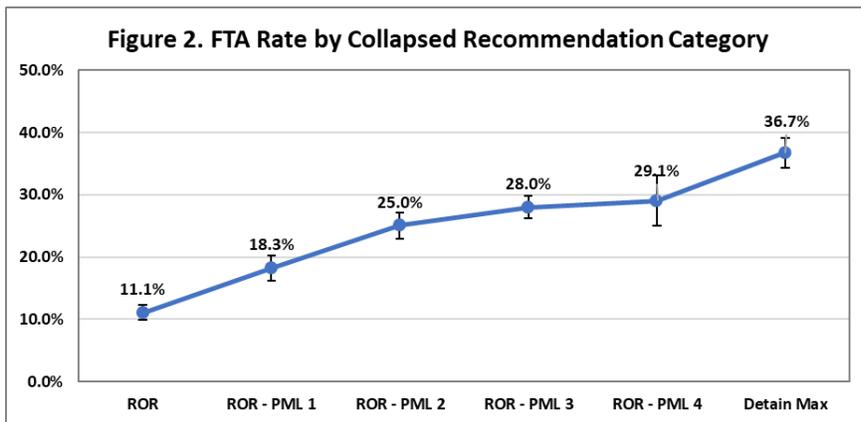
New Criminal Activity Rates

Court data was reviewed to determine if there was an NCA during the pretrial period of the case. As noted previously, City and County ordinances were not considered new criminal activity and the NCA Rate includes any violent criminal activity. The NCA Rate for all cases was 19.0% with a corresponding Public Safety Rate of 81.0%. Table 17 displays the NCA Rates for each recommendation category. The NCA Rate was highest in (U) Detain/Max and (Y) Detain/Max categories, at 33.1% and 32.8%, respectively, followed by (V) Detain/Max (31.1%). The lowest NCA Rates were observed in the (A) ROR, 8.3%, and (D) ROR, 8.6%, categories.

Table 17. NCA Rates by PSA Recommendation													
		New Criminal Activity Scale											
		1		2		3		4		5		6	
		NCAs	NCA Rate	NCAs	NCA Rate	NCAs	NCA Rate	NCAs	NCA Rate	NCAs	NCA Rate	NCAs	NCA Rate
FTA Scale	1	(A) ROR 105	8.3%	(B) ROR 92	12.9%								
	2	(C) ROR 19	8.4%	(D) ROR 55	8.6%	(E) ROR PML 1 101	16.7%	(F) ROR PML 3 72	21.7%	(G) ROR PML 4[11] 2	100.0%		
	3			(H) ROR PML 1 72	11.7%	(I) ROR PML 2 181	17.1%	(J) ROR PML 3 173	19.6%	(K) ROR PML 4 23	27.4%	(L) Detain or Max Conditions[9] 0	0.0%
	4			(M) ROR PML 1 34	17.6%	(N) ROR PML 2 91	20.5%	(O) ROR PML 3 140	26.7%	(P) ROR PML 4 112	28.1%	(Q) Detain or Max Conditions 21	28.4%
	5			(R) ROR PML 2 7	35.0%	(S) ROR PML 2 45	24.5%	(T) ROR PML 3 139	26.1%	(U) Detain or Max Conditions 168	33.1%	(V) Detain or Max Conditions 80	31.1%
	6							(W) Detain or Max Conditions 41	28.5%	(X) Detain or Max Conditions 38	24.4%	(Y) Detain or Max Conditions 139	32.8%

Figure 3 displays the NCA Rate with the associated 95% confidence intervals by the NCA score. As the NCA score increased, the NCA Rate increased and the Public Safety Rate decreased. For those who scored a 1 on the NCA scale, their NCA Rate was 8.3%, followed by 11.9% for NCA 2, 18.2% for NCA 3, 23.4% for NCA 4, and 29.9% for NCA 5. For individuals who had the highest NCA score of 6, the NCA Rate increased to 31.6%. The change from NCA 5 to NCA 6 was smaller than the change between any other two consecutive scores (1.7%). The Public Safety Rate ranged from 91.7% for cases with an NCA 1 down to 68.4% for cases with an NCA 6.


Figure 3. NCA Rate by NCA Score
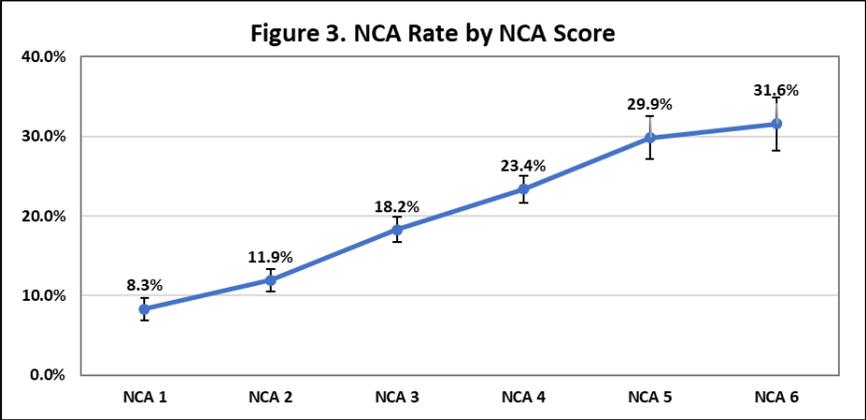
Figure 4 displays the NCA Rate by collapsed recommendation category. Like the pattern observed in the FTA Rate, the NCA Rate increased and the Public Safety Rate decreased as the recommendation category became more restrictive. Individuals in the ROR category had an NCA Rate of 9.5%. For ROR – PML 1, the NCA Rate increased to 14.6%, then 19.0% for ROR – PML 2, 23.0% for ROR – PML 3, 28.2% for ROR – PML 4, and 31.1% for individuals in the Detain/Max category. The Public Safety Rate ranged from 90.5% for cases with an ROR recommendation down to 68.9% for cases with a Detain/Max recommendation.

---

[11] Category G and L have few to no cases so a representative NCA Rate cannot be reliably calculated.

Figure 4. NCA Rate by Collapsed Recommendation Category
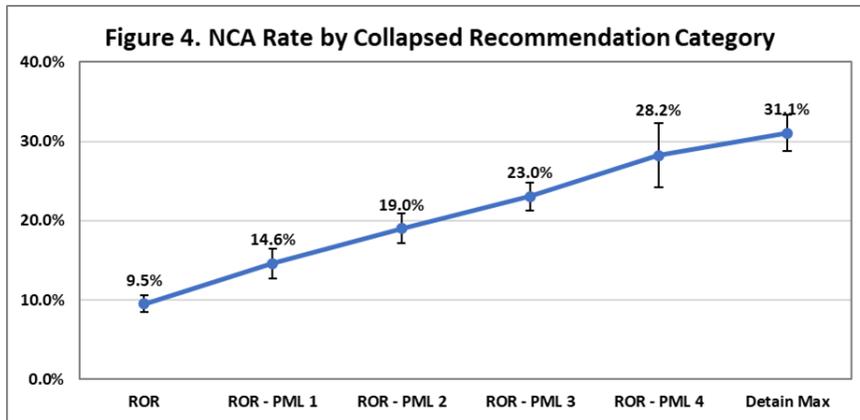
New Violent Criminal Activity Rates

Statutes were identified as violent by the Bernalillo County PSA Implementation Team with Arnold Ventures. These offenses include murder, kidnapping, robbery, assault, sex offenses such as rape and sexual assault, and conspiracy to commit these offenses. The NVCA Rate for all cases was 4.7% with 95.3% of individuals not having a new violent offense. Table 18 displays the NVCA Rate by PSA recommendation category. The highest NVCA Rate was observed in the (Q) Detain or Max Conditions category, at 13.5 % (10 out of 74), followed by (K) ROR – PML 4 at 10.7% (9 out of 84). The lowest NVCA Rate was in the (A), (C), and (H) categories, at 2.5%, 2.7%, and 2.8%.

Table 18. NVCA Rates by PSA Recommendation Category														
New Criminal Activity Scale														
		NCA 1		**NCA 2**		**NCA 3**		**NCA 4**		**NCA 5**		**NCA 6**		
		Count	NVCA Rate	Count	NVCA Rate	Count	NVCA Rate	Count	NVCA Rate	Count	NVCA Rate	Count	NVCA Rate
FTA Scale	**FTA 1**	(A) ROR 32	2.5%	(B) ROR 24	3.4%								
	FTA 2	(C) ROR 6	2.7%	(D) ROR 22	3.4%	(E) ROR PML 1 18	3.0%	(F) ROR PML 3 28	8.4%	(G) ROR PML 4[6] 0	0.02%		
	FTA 3			(H) ROR PML 1 17	2.8%	(I) ROR PML 2 47	4.4%	(J) ROR PML 3 45	5.1%	(K) ROR PML 4 9	10.7%	(L) Detain or Max Conditions 0	0.0%
	FTA 4			(M) ROR PML 1 9	4.7%	(N) ROR PML 2 24	5.4%	(O) ROR PML 3 23	4.4%	(P) ROR PML 4 27	6.8%	(Q) Detain or Max Conditions 10	13.5%
	FTA 5			(R) ROR PML 2 1	5.0%	(S) ROR PML 2 9	4.9%	(T) ROR PML 3 23	4.3%	(U) Detain or Max Conditions 37	7.3%	(V) Detain or Max Conditions 16	6.2%
	FTA 6							(W) Detain or Max Conditions 5	3.5%	(X) Detain or Max Conditions 9	5.8%	(Y) Detain or Max Conditions 38	9.0%

New Violent Criminal Activity Flag

While the NVCA flag is related to the PSA recommendation category because it is created using some of the same factor, it can indicate a higher risk of new criminal activity being violent. Judges may use this indicator to determine their final COR. For the 10,289 individuals in the sample, 16.0% or 1,651, had the NVCA flag.

The NVCA Rate for cases where there was an NVCA flag on the PSA assessment was over twice as high as cases where there was no flag (see Figure 5). The NVCA Rate for cases with no flag was 3.9% and for the rate for those with a flag was 8.7%, 123.1% higher than those with no NVCA flag.

Figure 5. NVCA Rate by NVCA Flag
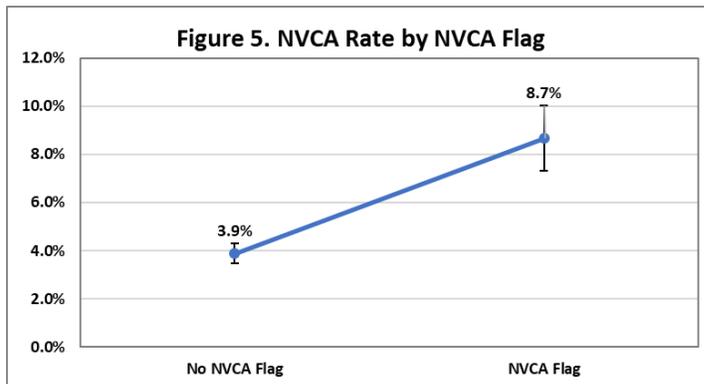
New Criminal Activity and Charge Details

Of the cases in the sample, 1,950 had an NCA during their pretrial period. The new criminal activity was reviewed and the data for the top three related charges were collected and assigned a charge level and charge category. The highest charge from those three were selected and were recorded as either a 1st degree felony (F1), a 2nd degree felony (F2), a 3rd degree felony (F3), a 4th degree felony (F4), misdemeanor (MD), or petty misdemeanor (PM). Next, the charge categories were coded into one of six crime type categories: violent, drug, property, DWI, and public order or other crimes. The highest charge was selected based on the highest charge level followed by the highest charge category (see Table 19). In general, the majority of new criminal activity was 4th degree felonies and misdemeanors in the property, drug, and violence categories. These categories of charges accounted for 1,291 (66.2%) of the 1,950 cases with new criminal activity. Nearly half (42.3%) of all NCAs were property charges, violent charges made up 22.8% of all NCAs, and drug charges were 23.6%.

Of the NCAs, nearly half (47.6% or 928) had the lowest level of a felony, F4. Misdemeanor charges and petty misdemeanor charges comprised slightly more than a third of all NCAs (25.8% and 10.3%, respectively). F3 accounted for 9.8% of all NCA. Finally, 5.8% (114) of the NCA cases had an F2 as the highest charge and 0.7% (13) had an F1 as the highest charge.

Table 19. NCA Highest Charge Level and Category							
	Violent	Drug	Property	DWI	Public Order/Other	*Total*	*Percent of all NCAs*
F1	10	3	0	0	0	13	0.7%
F2	42	46	23	0	3	114	5.8%
F3	99	20	69	1	1	190	9.8%
F4	134	341	413	2	38	928	47.6%
MD	121	53	229	10	91	504	25.8%
PM	45	9	77	12	58	201	10.3%
Total	475	491	781	25	178	1,950	100.0%
Percent of All NCAs	24.4%	25.2%	40.1%	1.3%	9.1%	100.0%	

The highest charge level was compared to the collapsed PSA recommendation category (see Table 20). There were few F1s (between 0.2% and 1.5% of cases) and F2s (between 4.1% and 10.2% of cases) across the recommendation categories. F3 was also a smaller category, accounting for between 8.0% and 11.4% of cases and petty misdemeanors was small as well, between 8.8% and 12.1% of cases by recommendation category. The largest two charge levels across the release types were misdemeanors (between 19.0% and 31.0% of cases) and F4s, which accounted for between 40.2% and 51.1% of cases.

Table 20. NCA Charge Level by Collapsed Recommendation Category													
	F1		F2		F3		F4		MD		PM		Total
ROR	4	1.5%	16	5.9%	31	11.4%	109	40.2%	84	31.0%	27	10.0%	271
ROR - PML 1	2	1.0%	11	5.3%	17	8.2%	90	43.5%	62	30.0%	25	12.1%	207
ROR - PML 2	2	0.6%	20	6.2%	37	11.4%	153	47.2%	78	24.1%	34	10.5%	324
ROR - PML 3	1	0.2%	33	6.3%	53	10.1%	268	51.1%	121	23.1%	48	9.2%	524
ROR - PML 4	2	1.5%	14	10.2%	13	9.5%	70	51.1%	26	19.0%	12	8.8%	137
Detain/Max	2	0.4%	20	4.1%	39	8.0%	238	48.9%	133	27.3%	55	11.3%	487
Total	13	0.7%	114	5.8%	190	9.7%	928	47.6%	504	25.8%	201	10.3%	1,950

Table 21 includes a comparison between the source case charge level – that is the charge level of the case for which the individual was assessed – and the charge level, if any, of the new criminal activity. Nearly half of the cases (47.5% or 926) of the cases with an NCA had a charge level lower than the case that they were assessed on. This is an important finding as it indicates that while 19.0% of individuals have an NCA during the pretrial period, it is most often a lower level charge or an equivalent level charge (39.3% or 765).

Table 21. Source Charge Levels and NCA Charge Level Comparisons							
	NCA Charges Lower than Source Case		NCA Charges Same as Source Case		NCA Charges Higher than Source Case		Total
Source Charge Level	Count	Percent	Count	Percent	Count	Percent	
F1	12	100.0%	0	0.0%	n/a	n/a	12
F2	127	89.4%	13	9.2%	2	1.4%	142
F3	270	81.8%	36	10.9%	24	7.3%	330
F4	517	35.2%	716	48.9%	233	15.9%	1,466
Total	926	47.5%	765	39.3%	259	13.3%	1,950

Predictive Validity of the PSA

The predictive validity of the PSA was assessed using Area Under the Curve (AUC) of the Receiver Operator Characteristics (ROC) estimates. The AUC is a general measure of a tool's accuracy often used in risk assessment, medical diagnosis, image recognition, and other fields. The Area Under the Receiver-Operator Characteristic (ROC) curve, is a measure of how the sensitivity and specificity of a test—the extent to which it correctly identifies positive or negative subjects—varies with the threshold we use to label a subject positive or negative.

The AUC score ranges from 0.0 to 1.0, and represents the degree to which the PSA creates more accurate estimates for pretrial success than random chance alone. A score of 0.5 is the equivalent of random chance and higher scores represent higher levels of predictive validity. A higher AUC indicates a higher probability that the tool will assign a higher risk score to an individual who will fail than to an individual who will not fail. An AUC of 0.65, for instance, means that the tool would put these two individuals in the correct order 65% of the time (treating ties as correct half the time).

In applications such as medical diagnosis, a test is considered acceptable if its AUC is between 0.7 and 0.8, and excellent if it is at least 0.8. Since human behavior is difficult to predict, the standards in risk assessment are lower. In a 2013 review of research on instruments used in the United States to assess the risk of recidivism, Desmarais and Singh used the following AUC score categories to rank the predictive validity: less than .55 was considered poor, .55 to .63 was fair, .64 to .71 was good, and .71 to 1.00 was considered excellent (p. 12).

Overall Predictive Validity

The AUC ROC estimates were calculated for the FTA, NCA, and NVCA. For the FTA, the AUC score was 0.641 (see Table 22). The AUC score was slightly lower for NCA (0.638) and NVCA (0.572). Based on the categories proposed by Desmarais and Singh (2013), the predictive utility of the PSA for FTA and NCA was considered good and the predictive validity of the PSA for NVCA was fair[12]. These scores indicate that for all three outcomes, the PSA provides more accurate estimates for pretrial success than random chance alone.

Table 22. Area Under the Curve Receiver Operator Characteristics	
	AUC Score +/- CI (N=10,289)
FTA	0.641 +/- .012
NCA	0.638 +/- .013
NVCA	0.572 +/- .028

---

[12] Note, while this score is lower than other validation research, the AUC is calculated elsewhere using the raw violent score which ranged from 0 to 7. The defendant is given a flag or not, so a dichotomized score is used and presented to the judge for decision-making. This score is what was used for AUC ROC analysis.

Outcome Measures by Race and Gender

The outcome measures were calculated and compared by race. While there was some variation in rates, the overall findings demonstrated that differences in the FTA, NCA, and NVCA rates by race was not statistically significant.

The FTA Rate was 21.6% for cases with a White defendant (see Figure 6). For cases with a Hispanic defendant the FTA Rate was 23.4% and the FTA Rate was slightly higher if the defendant was Native American (24.3%) or Black (24.3%). These rates correspond with Appearance Rates of 78.4% for cases with White defendants, 76.6% for Hispanic defendants, and 75.7% for Native American and Black defendants. The NCA Rate was lower for cases with a Native American defendant (14.5%) than any other race by several percent, with the next lowest NCA in cases with a Black defendant (18.7%), followed by White defendants (19.0%), and Hispanic defendants (19.6%). These rates correspond with Public Safety Rates of 85.5% for cases with Native American defendants, 81.3% for Black defendants, 81.0% for White defendants, and 80.4% for Hispanic defendants. The NVCA Rate was lowest for cases with a White defendant (4.0%) or a Hispanic defendant (4.8%) and highest for cases with a Native American defendant (6.1%) or a Black defendant (6.3%). A chi-square (X^2) test of independence was performed to determine if there was a correlation between race and the outcome results[13]. The results indicated there was no significant association between race and FTA [X^2 (3, N=9,968) =5.387, p=.146]. There was no significant association between race and NCA [X^2 (3, N=9,968) =11.184, p=.011] or NVCA [X^2 (4, N=9,968) =11.287, p=.010][14].

Figure 6. Outcome Measures by Race Category
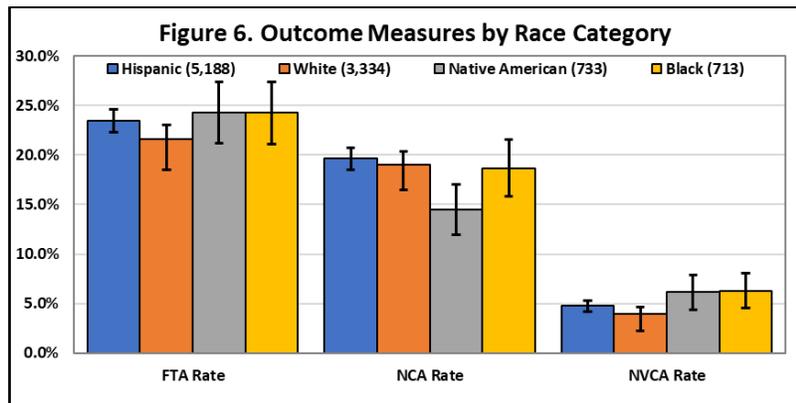
The outcome measures were calculated by gender (see Figure 7). By gender, there was little variation in the FTA Rates and this variation was not significant. The correlation between gender and the NCA Rate was statistically significant and the effect size was large and the correlation with the NVCA Rate was also significant with a medium-large effect size.

The FTA Rate was similar for male and female individuals (22.8% and 23.3%) and this difference was not statistically significant. The Appearance Rate for male individuals was 77.2% and 76.7% for female individuals. The NCA Rate was nearly 5% higher for male individuals than female individuals (20.2% compared to 15.4%). This corresponds with a Public Safety Rate of 79.8% for male individuals and 84.6% for female individuals. Based on the chi-square (X^2) test of independence, the association between

---

[13] Excluding the Other race category, N=9,968.

[14] Although significance would have been achieved for NCA and NVCA if a lower significance level had been used, the effect size was very small (Cramer's V=.033 and .034, respectively).

gender and NCA was both significant [X^2 (1, *N*=10,289) =29.212, p<.001] and the effect size was large (Cramer's V=.053). The NVCA Rate was 2.2% higher for male than female individuals (5.2% compared to 3.0%). This association was significant [X^2 (1, *N*=10,289) =21.263, p<.001] with a medium to large effect size (Cramer's V=.045).
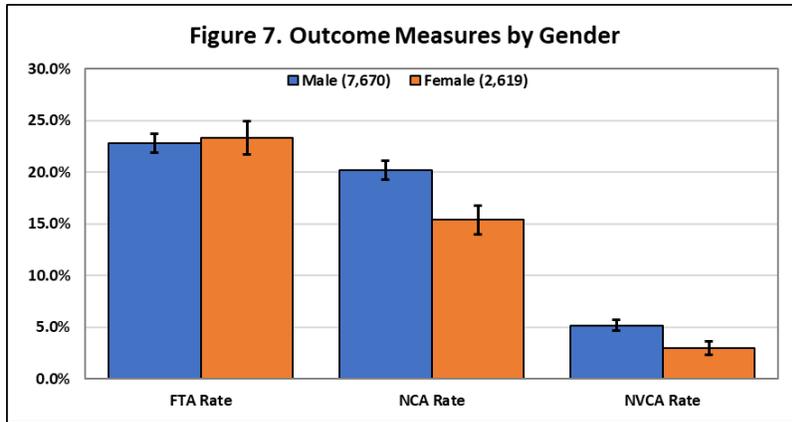
Overall, differences in outcomes occur mainly in the NCA and NVCA Rate. Where these differences occurred by race, the effect size was small, however by gender these differences were both significant and large.

Predictive Validity by Race and Gender

The AUC scores by race were calculated for each category of outcomes and compared to the scores for the overall population. The AUC scores for FTA and NCA among cases where the defendant was Hispanic, White, Native American, or Black are shown in Table 23. There was some variation between groups, but these variations were not statistically significant. Thus, the PSA has similar predictive validity within each demographic group, i.e., a similar ability to assign lower scores to defendants who will succeed than to defendants, in the same demographic group, who will have an FTA or NCA outcome.

According to Desmarais and Singh (2013), the AUC scores by race would be considered fair for FTA and NCA with a few exceptions. For FTA for White or Native American individuals (0.628 and 0.614 respectively) and NCA for Hispanic individuals (0.630) the AUC would be considered fair rather than good. The AUC scores by race for NVCA were considered fair except for cases where the defendant was Black, for which the predictive validity for NVCA was poor (0.543). While there were differences across each group for all three categories, these differences were not statistically significant and were less than two standard deviations away from the AUC scores from the entire population.

Table 23. Area Under the Curve Receiver Operator Characteristics Scores and CIs by Race					
	Hispanic (N=5,188)	White (N=3,334)	Native American (N=733)	Black (N=713)	Overall (N=10,289)
FTA	0.642 +/- 0.017	0.628 +/- 0.022	0.614 +/- 0.048	0.683 +/- 0.045	0.641 +/- .012
NCA	0.630 +/- 0.018	0.636 +/- 0.024	0.645 +/- 0.056	0.661 +/- 0.050	0.638 +/- .013
NVCA	0.558 +/- 0.039	0.576 +/- 0.050	0.619 +/- 0.090	0.543 +/- 0.090	0.572 +/- .028

*p<.001

The AUC scores by gender were calculated for the FTA, NCA, and NVCA outcomes (see Table 24). The differences by gender in the AUC scores were not statistically significant.

For cases where the defendant was male, the AUC score for FTA was 0.643 compared to 0.638. These scores are very similar and would be considered good. Differences were larger for both the NCA and NVCA, though these differences were still not statistically significant. For cases with a male defendant the AUC score for NCA was 0.639 (good) while the AUC score for females was 0.623 (fair). For NVCA, cases with a male defendant had an AUC score of 0.569 (fair) and 0.568 for female defendants (fair).

Table 24. Area Under the Curve Receiver Operator Characteristics Score by Gender		
	Male (N=7,670)	Female (N=2,619)
FTA	0.643 +/- 0.014	0.638 +/- 0.025
NCA	0.639 +/- 0.015	0.623 +/- 0.029
NVCA	0.569 +/- 0.031	0.568 +/- 0.069

*p<.001

The AUC scores indicate that the PSA provides fair to good predictive validity for individuals regardless of race or gender for all outcome measures. This is an indicator that the PSA provides a more accurate assessment than random selection alone.

PSA Scores and Outcomes by Race and Gender

The FTA Rate was compared across race categories for each related score category (see Figure 8). There was variation across race categories for some scores, but this variation was not statistically significant based on the results of a chi-square test.

The FTA Rate for individuals with an FTA score of 1 ranged from 7.4% for cases with a Black defendant, and up to 14.7% for cases with Native American defendants (see Figure 8). The difference among individuals was smallest for cases where the FTA score was 2 (between 14.6% for Hispanic individuals and 19.0% for Native American individuals) or 3 (22.1% for Black individuals and 26.0% for Native American individuals). For an FTA score of 4, the FTA Rate ranged from 18.1% for Native American individuals and 33.3% for Black individuals and for an FTA score of 5, 30.0% for White individuals up to 41.2% for Native American individuals. The category with the largest range was FTA 6, from 36.8% for White individuals and 53.7% for Black individuals.

In addition to the chi-square test, a Neyman smooth test (see e.g. Ghosh & Bera, 2001) was conducted to combine significance levels across the 6 score categories to see if one group has failure rates that are systematically higher or lower rather than in an individual category: that is, if defendants in this group have rates different from other defendants with the same score (technically, if there is a difference when conditioned on score). This test was used to combine results across the six FTA score categories. We found no statistically significant differences in the FTA by race using this method.

Figure 8. FTA Rates by Race and FTA Score
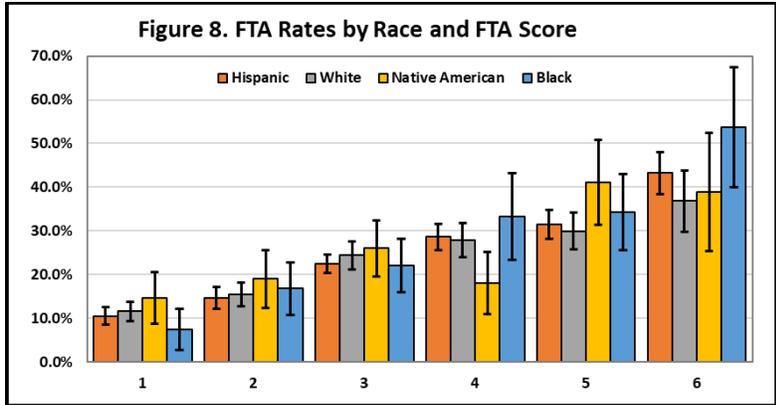
The NCA Rate was compared across race categories for each related score category (see Table 29). There was variation across race categories for some scores, but this variation was not statistically significant. There was similar variation in the NCA Rate by race and the NCA score, although the range was smaller. The NCA Rate for cases with an NCA 1 ranged from 7.0% among cases with a Black defendant and up to 9.6% for cases with a White defendant (see Figure 9).

The NCA Rate among cases where the NCA score was 2 and the defendant was Native American was 6.5% and the rate was twice as high among White defendants, 13.0%. There was a similar gap in the NCA Rate between Native American and Hispanic individuals with an NCA score of 3 (12.5% and 18.6%). Those with an NCA score of 4 ranged from an NCA Rate of 18.2% for Black individuals up to 24.9% for White individuals. The range for NCA 5 and 6 were larger, ranging from 23.4% for Native American individuals up to 36.1% for Black individuals (NCA 5) and 23.5% for Native American individuals and 34.3% for White individuals (NCA 6). Based on the results of the chi-square test, the association between race, NCA score, and NCA was not statistically significant.

The Neyman smooth test was used to combine significance levels across the 6 NCA score categories. We found no racial differences that met our threshold p-value of 0.001 for statistical significance. However, we did find weaker evidence, with p-value 0.002, that NCA Rates for Native American defendants are lower than the overall population when conditioned on NCA score. We suggest studying this potential difference in the future as more data becomes available.

Figure 9. NCA Rates by Race and NCA Score
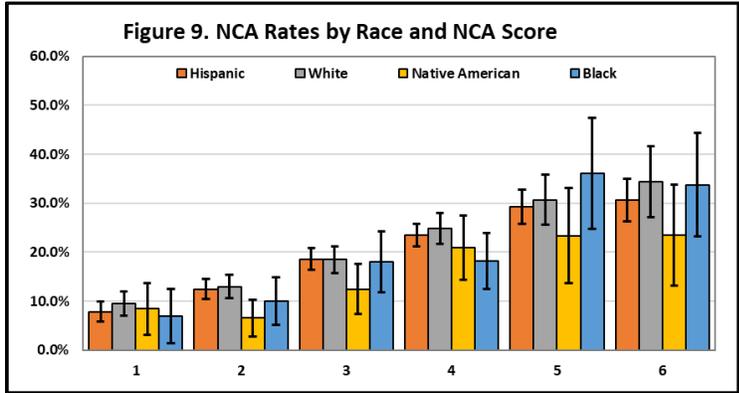
The NVCA Rate was compared across race categories for the NVCA flag (see Figure 10). There was variation across race categories for some scores, but this variation was not statistically significant[15].

There was a small amount of variation in the NVCA Rate in race by whether the individual had the NVCA flag. The NVCA Rate for those without the NVCA flag was lowest among cases where the individual was White (3.3%) and highest among Black individuals (5.6%) (see Table 30). Among those with an NVCA flag, the NVCA Rate was lowest among Hispanic individuals at 8.0% and highest among Native Americans at 11.3%. Based on the results of the chi-square test, the association between race, the NVCA flag, and NVCA was not statistically significant.

Figure 10. NVCA Rates by Race and NVCA Flag
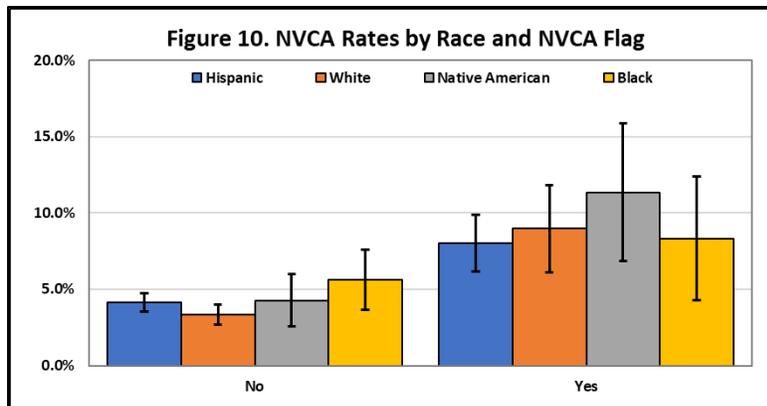
The FTA Rate was compared by gender for each related score category (see Figure 11). There was variation by gender for some scores, but this variation was not statistically significant.

By gender, the FTA Rate was higher for cases where the individual was a female for all FTA scores except FTA 4 where the female FTA Rate was 28.8% and the male FTA Rate was 3.0% lower at 25.8% (see Figure 11). In the remaining categories, the FTA Rate for females was higher by as little as 0.3% for FTA 5 to as much as 4.1% for FTA 6. Based on the results of the chi-square test, the association between gender, the FTA score, and the FTA was not statistically significant.

The Neyman smooth test was used to combine significance levels across the 6 FTA scores. This test was used to combine results across the six FTA score categories and it was determined that there was no significant difference in the FTA by gender across the FTA scores.

---

[15] The Neyman smooth test is not used in this study for the NVCA as the NVCA is a dichotomous score and it makes little difference to the results.

Figure 11. FTA Rates by Gender and FTA Score
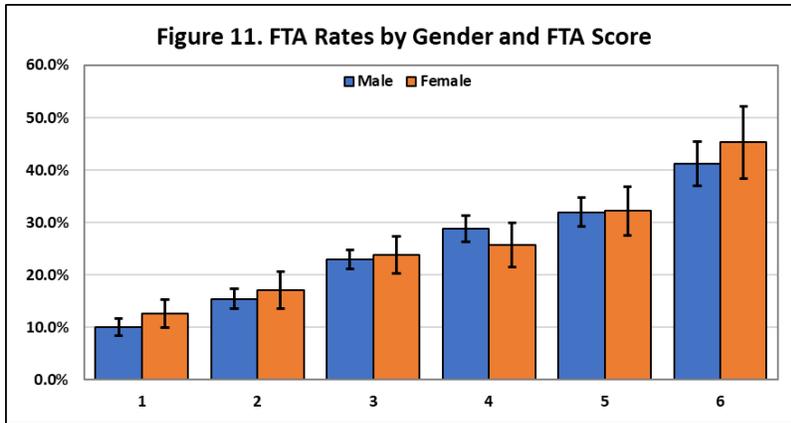
The NCA Rate was compared by gender for each related score category (see Figure 12). While the male individuals consistently had higher NCA Rates than females, these differences were not statistically significant. The NCA Rates are most similar by gender for NCA 1, 2, and 3 with differences ranging from 0.6% and 1.7% (see Figure 12). The differences in the NCA Rate was much higher for NCA 4 (4.2%), NCA 5 (9.5%), and NCA 6 (5.2%). Based on the results of the chi-square test, the association between gender, the NCA score, and the NCA was not statistically significant[16].

The Neyman smooth test was used to combine significance levels across the 6 NCA scores. While the results of the chi-square test showed that the association between gender and the NCA scores was not statistically significant in any one score category, combining the six score categories with the Neyman smooth test, we find that female NCA Rates are significantly lower than male NCA Rates overall across NCA scores, with p=0.001. That is, female defendants have NCA outcomes less often than male defendants with the same NCA score.

Figure 12. NCA Rates by Gender and NCA Score
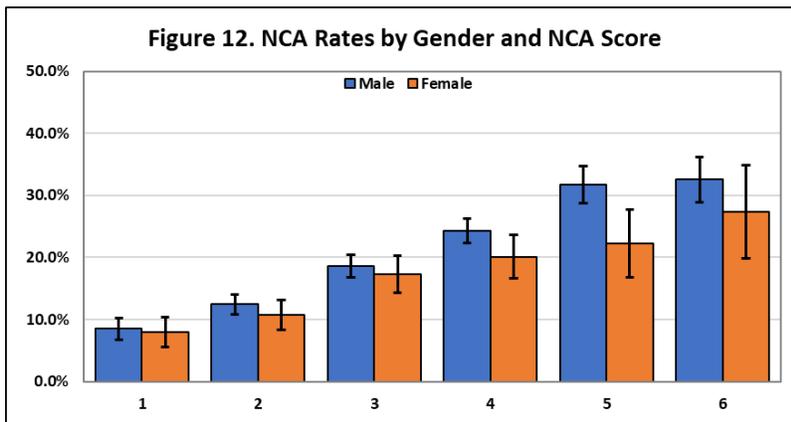
The NVCA Rate was compared by gender for the NVCA flag (see Figure 13). There was variation by gender that was statistically significant with a large effect size.

There was variation in the NVCA Rate by gender and the NVCA flag. The NVCA Rate for those cases without the NVCA flag where the defendant was male was 4.4%, 1.8% higher than the rate of 2.6% for

---

[16] Significance would have been achieved for NCA 4 and NCA 5 if a lower significant level had been used and the effect sizes would have been medium (Cramer's V=.040) and large (Cramer's V=.083), respectively.

females with no NVCA flags (see Figure 13). For those cases with the NVCA flag where the defendant was male, the NVCA Rate was 8.9% and for the cases with a female defendant the rate was 7.1%. Based on the results of the chi-square test, the association between gender, the NVCA flag, and the NVCA Rate was statistically significant for cases with no NVCA flag (p<.001) and of a medium effect size (Cramer's V=.042).

Figure 13. NVCA Rates by Gender and NVCA Flag
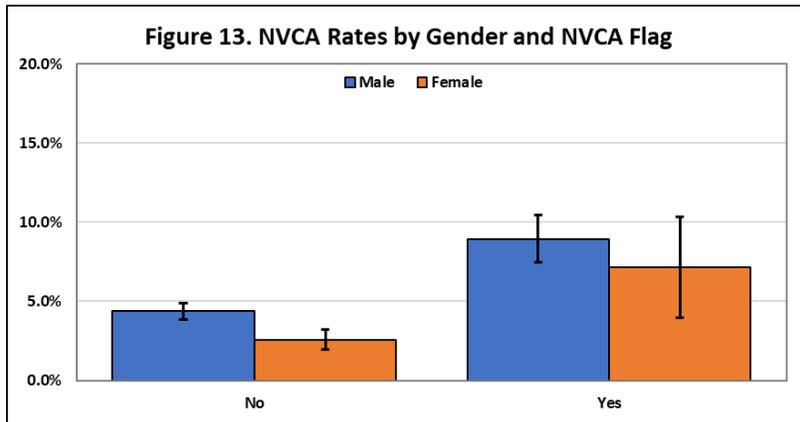
Overall, there were no significant differences in the AUC scores by race or by gender. There were also no statistically significant differences in the FTA Rates by race or gender. The differences in the NCA and NCVA Rates by race were not statistically significant, although we found weak evidence that Native American defendants may have lower NCA Rates, and future studies should examine this further. The only significant findings were the gender differences in NCA, and gender differences in NVCA Rates for cases where the NVCA flag is absent. Both of these have a large effect size.

Predictive Fairness by Race and Gender: Odds Ratio

In addition to the tests for disparities in FTA and NCA Rates, the data was also fit with logistic regression as done in some previous validation studies, including the Kentucky (DeMichele et al., 2018) and Kane County (Greiner et al., 2021) studies. Logistic regression produces an "odds ratio" which estimates the increase in the probability of failure whenever a score increases by 1, and a moderated regression can be used to determine if this odds ratio is significantly different between different demographic groups.

There are several methodological problems with this approach. First, the FTA and NCA scores are ordinal data rather than interval data: they are abstract categories produced by applying various thresholds to a raw score derived from the input factors, so there is little reason to assume that, say, going from 2 to 3 would increase failure rates by the same ratio as going from 4 to 5. Second, the parameters of the logistic function are good test statistics for fairness if the logistic model fits the data, but the logistic function doesn't fit the actual failure rates very well, especially in the higher score categories.

In contrast, the Neyman test used above allows us to test hypotheses like "across score categories, defendants in group A have higher/lower failure rates than defendants in group B with the same score" without assuming anything about how either group's rates increase as a function of the score. Nevertheless, we include the logistic regression as an additional analysis to make this study easier to compare with previous validation studies. The figures below show estimated failure rates using logistic regression, where the colored bands indicate 95% confidence intervals. The actual failure rates are shown for comparison.

Odds Ratios and Race

Tables 25 shows odds ratios with 95% confidence intervals for FTA, NCA, and NVCA by race. An odds ratio of 1.42, for instance, is an estimate that each increase in score corresponds to a 42% increase in the failure rate. In some instances, the odds ratio is much higher due the amount of change between the lower and higher scores. The confidence intervals all overlap, and showed no statistically significant differences. The confidence intervals are quite broad for the NVCA odds ratios but are narrower for the FTA and NCA odds ratios.

| Table 25. Odds Ratios and Confidence Intervals by Race						
	FTA Odds Ratio	95% Conf. Interval	NCA Odds Ratio	95% Conf. Interval	NVCA Odds Ratio	95% Conf. Interval
Hispanic	1.42	1.35–1.48	1.38	1.31–1.45	2.03	1.51–2.71
White	1.34	1.27–1.42	1.41	1.32–1.50	2.86	1.92–4.28
Native American	1.31	1.17–1.47	1.39	1.21–1.60	2.87	1.56–5.28
Black	1.55	1.38–1.75	1.48	1.30–1.69	1.52	0.80–2.90

For race, the fitted FTA Rates overlap significantly (see Figure 14). Visually, there is a suggestion that FTA Rates are higher for Black defendants at the higher score levels, but the actual failure rate for this group with FTA score 5 is well below the logistic fit, and neither the chi-square test or the Neyman smooth test found a statistically significant difference.

Figure 14. FTA Rates and Logistic Fits by Race
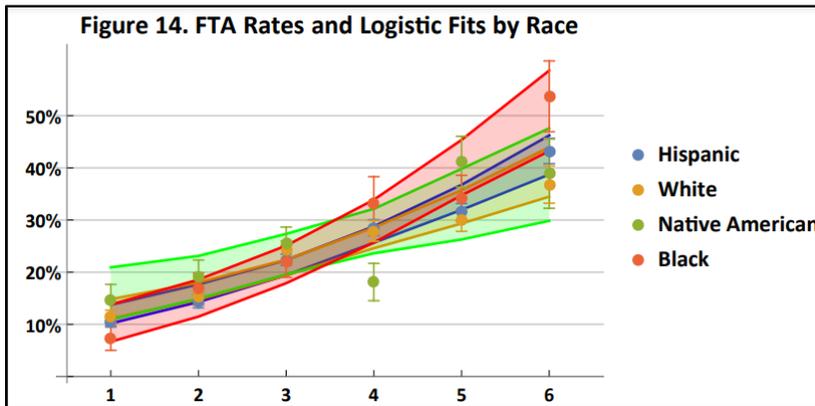
For NCA, the fitted probability for Native American individuals is noticeably below the other groups over several score categories (see Figure 15). While this visualization is not convincing statistical evidence on its own, it reproduces our finding in the previous section that these individuals have lower NCA Rates at significance level p=0.002.

Figure 15. NCA Rates and Logistic Fits by Race
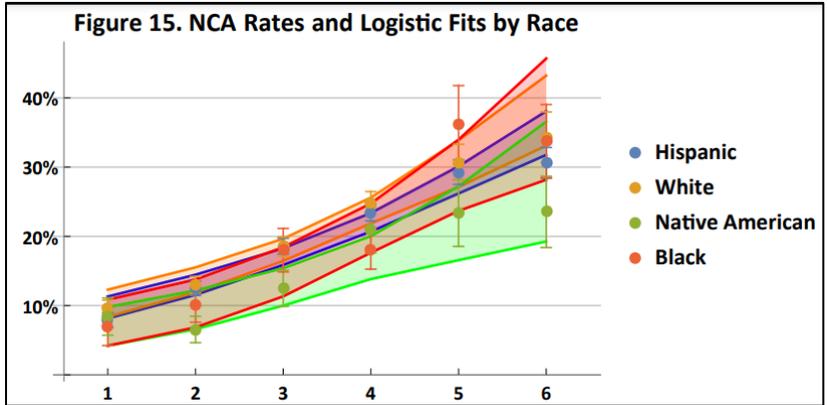
For NVCA, the 95% confidence intervals overlap considerably (Figure 16), reproducing our result from the previous section that there are no statistically significant racial differences in NVCA Rates, with or without the NVCA flag.


Figure 16. NVCA Rates and Logistic Fits by Race
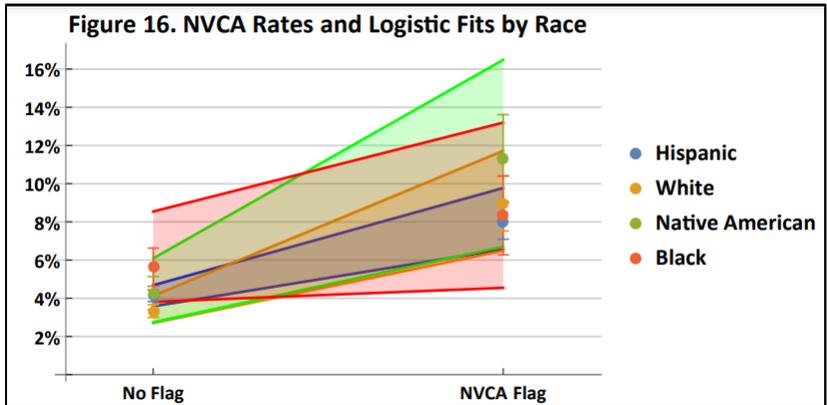
Odds Ratios and Gender

Table 26 shows odds ratios with 95% confidence intervals for FTA, NCA, and NVCA by gender. An odds ratio of 1.41, for instance, is an estimate that each increase in score corresponds to a 41% increase in the failure rate. In the case of the NVCA, the odds ratio is much higher due the amount of change between the lower and higher scores. The confidence intervals overlap, and there were no statistically significant differences by gender.

|	FTA Odds Ratio	95% Conf. Interval	NCA Odds Ratio	95% Conf. Interval	NVCA Odds Ratio	95% Conf. Interval
Table 26. Odds Ratios and Confidence Intervals by Gender						
Male	1.41	1.36–1.46	1.41	1.35–1.47	2.14	1.72–2.67
Female	1.37	1.29–1.45	1.34	1.25–1.45	2.91	1.69–5.00

For gender, the confidence intervals overlap for FTA (Figure 17), showing no statistically significant differences. This corresponds with the findings above on the differences in FTA Rates and gender.

Figure 17. FTA Rates and Logistic Fits by Gender
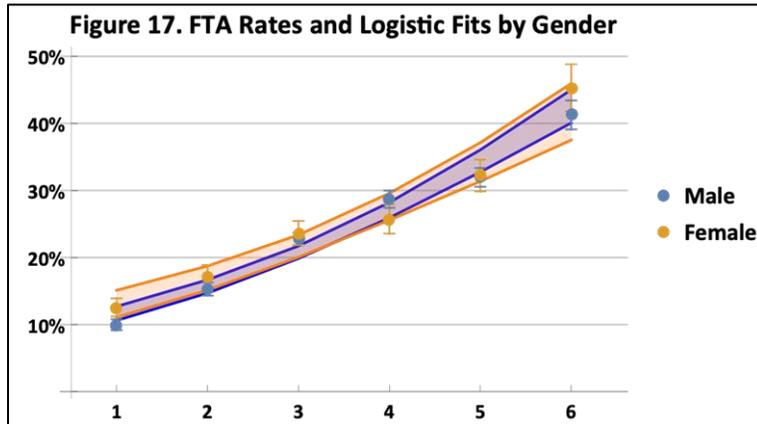
The fitted confidence intervals for NCA are distinct (Figure 18), with the logistic fit consistently lower for female individuals. While this visualization is not statistical evidence on its own, this reproduces our earlier result that women have lower NCA Rates than men.

Figure 18. NCA Rates and Logistic Fits by Gender
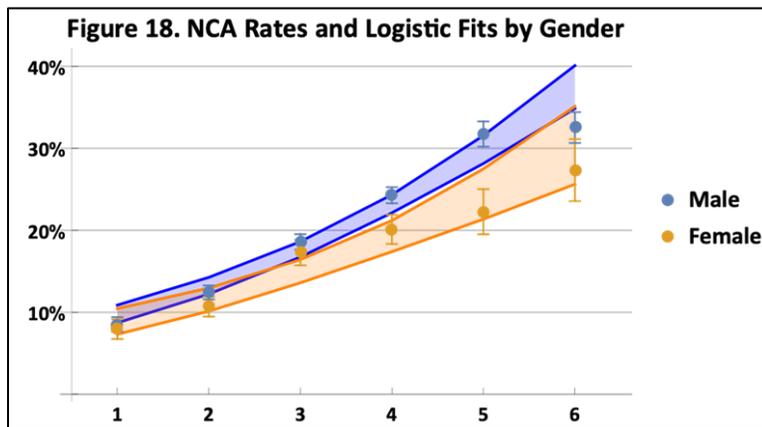
For NVCA, Figure 19 shows that the 95% confidence intervals are well-separated by gender for individuals when the flag is absent and overlap when the flag is present. This reproduces the result of the previous section that NVCA Rates are significantly lower for women than for men in cases without the flag, but are statistically indistinguishable for cases with the flag.

Figure 19. NVCA Rates and Logistic Fits by Gender
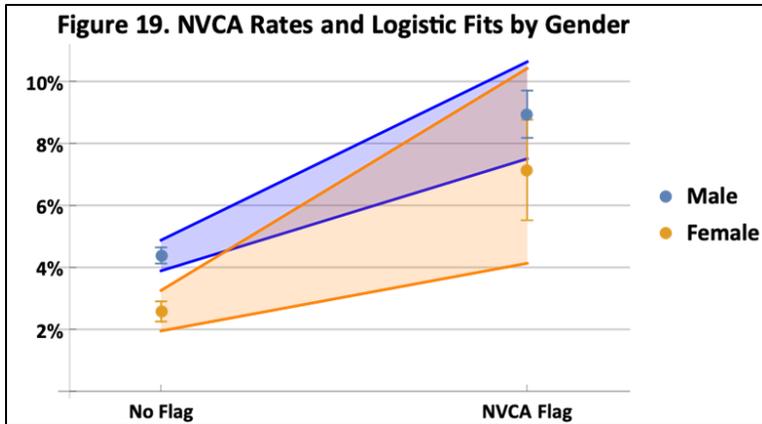
Adherence

Adherence describes the degree to which the COR as recommended by the PSA correspond to the COR ordered by the judge. For example, if an individual's PSA recommendation category was ROR on the DMF, and the judge ordered the individual to be released on his own recognizance, the conditions adhered to the PSA release recommendation. When the judge added more conditions, such as pretrial supervision (PTS), it was considered more restrictive. If the conditions set have fewer conditions than the PSA recommendation these were considered less restrictive. Only cases that had COR ordered were included in the adherence review. Therefore, cases for which there was a preventive detention motion (PTD) filed would not qualify as there are no conditions being set by the judge at the initial hearing[17]. Shown in Table 27 below, the 10,289 cases from the outcome sample were reduced by the number of cases in which there was a PTD motion or no conditions of release. There were 1,167 cases excluded, leaving 9,122 cases for adherence review.

Table 27. Exclusions from Adherence Sample			
Exclusion Criteria	BCMC	SJDC	Total
Preventive Detention Motion	1,158	2	1,160
No COR, No Bond Hold for PTD on Other Case	0	7	7
Total Excluded	1,158	9	1,167

Table 28 displays the adherence or deviation by the collapsed recommendation categories. Overall, the conditions set cases adhered to the PSA recommendation in 76.5% of cases. Conditions were less restrictive in 8.6% of cases and more restrictive in 14.9% of cases.

Of the 2,461 cases where the individual was recommended ROR, the judge ordered more restrictive conditions 22.3% of the time. For the various levels of ROR with PML, the adherence ranged from 67.9% for ROR – PML 1 up to 80.8% for ROR – PML 3. Less restrictive conditions than the recommendation that was ordered at a decreasing rate from 27.6% for PML 1, 13.4% to PML 2, 7.5% for PML 3, and 5.4% for PML 4. Conversely, the more restrictive conditions were assigned at an increasing rate, starting at 4.4% for PML 1, 7.9% for PML 2, 11.7% for PML 3, and 17.8% for PML 4. Of the cases where the individual had

---

[17] Preventive detention motions are typically filed during the BCMC portion of the case. The decision regarding this motion cannot be made by a BCMC judge and the defendant in the interim is put on a temporary hold pending an evidentiary hearing in SJDC. The BCMC judge does not have the ability to set conditions of release for defendants in these circumstances.

Detain/Max recommended, the judge ordered more conditions 22.5% of the time and less restrictive conditions 3.9% of the time.

Table 28. Adherence or Deviation by Collapsed Recommendation Category					
PSA Collapsed Category	Total	Percent of All Adherence Cases	Less Restrictive	Adherence	More Restrictive
ROR	2,461	27.0%	-	77.7%	22.3%
ROR - PML 1	1,260	13.8%	27.6%	67.9%	4.4%
ROR - PML 2	1,549	17.0%	13.4%	78.8%	7.9%
ROR - PML 3	2,007	22.0%	7.5%	80.8%	11.7%
ROR - PML 4	428	4.7%	5.4%	76.9%	17.8%
Detain/Max	1,417	15.5%	3.9%	73.6%	22.5%
Total	*9,122*	*100.0%*	*8.6%*	*76.5%*	*14.9%*

The FTA Rates were calculated by adherence or deviation from the recommendation (see Figure 20). The FTA Rate was lowest (17.9%) for cases where conditions were less restrictive and next lowest for cases where the recommendation and COR corresponded (22.2%). The corresponding Appearance Rates were therefore highest when conditions were less restrictive (82.1%) and lower when conditions corresponded with the recommendation (77.8%). The FTA Rate was highest when the COR were more restrictive than the release recommendation at 27.9% with a corresponding Appearance Rate of 72.1%. Based on a chi-square test, there was a significant association between adherence and FTA [X^2(2, N=9,122) =32.171, p=<.001], although the size of this effect was relatively small (Cramer's V=.059).

Figure 20. FTA Rate by Adherence or Deviation
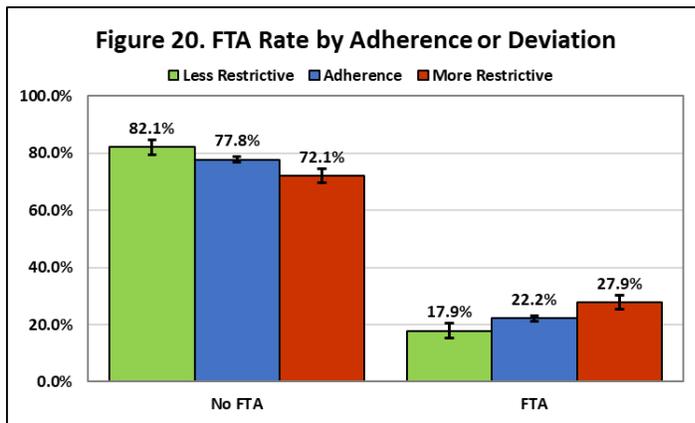
Next, the NCA Rates were calculated by adherence or deviation, displayed in Figure 21. Similar to the FTA Rates, the failure rates were highest when the judge ordered more restrictive conditions than recommended on the PSA. The NCA Rate for those with less restrictive conditions was 15.3% and 18.1% when the conditions adhered to the recommendation. The corresponding Public Safety Rates were therefore highest when conditions were less restrictive (84.7%) and lower when conditions corresponded with the recommendation (81.9%). The NCA Rate increased to 21.6% for those with more restrictive conditions with a corresponding Public Safety Rate of 78.4%. Based on a chi-square test, there was a significant association between adherence and NCA [X^2(2, N=9,122) =14.571, p=<.001], although the size of this effect was relatively small (Cramer's V=.040).

Figure 21. NCA Rate by Adherence or Deviation
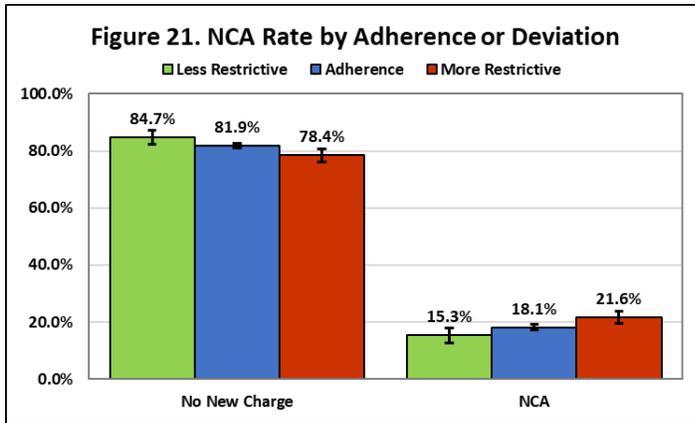
Figure 22 shows the NVCA Rates were calculated by adherence or deviation. Similar to the FTA and NCA Rates, the failure rates were highest when the judge ordered more restrictive conditions than recommended on the PSA, though with very little variation. The NVCA Rate for those with less restrictive conditions was 3.4% and 4.3% when the conditions adhered to the recommendation. In terms of pretrial success this means, 96.6% of individuals with less restrictive conditions did not have a new violent offense and 95.7% when conditions followed the PSA recommendations did not have a new violent offense. The NCA Rate increased to 4.9% for those with more restrictive conditions. There was no significant association between adherence and NVCA [X^2 (2, N=9,122) =2.464, p=.292].

Figure 22. NVCA Rate by Adherence or Deviation
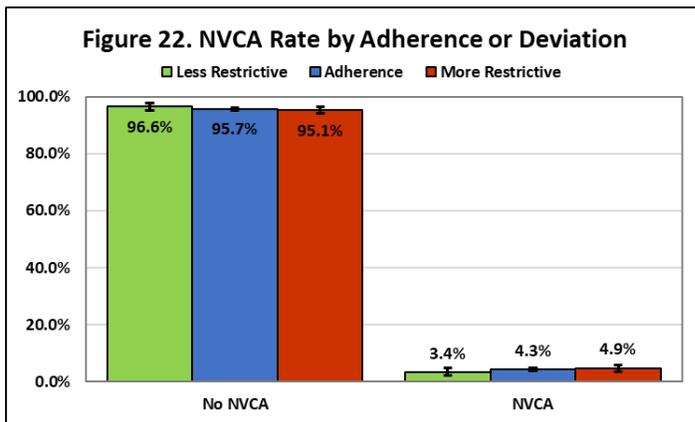
Discussion

The current study is the first validation of the PSA in Bernalillo County since its use began in 2017. Bernalillo County, like all places, is unique—not only in its composition and geographical location, but in the use of the PSA only for felony cases. In addition, in 2017, preventive detention motions began, as did bail reform, and in 2018, BCMC preliminary hearings began which were the only source of FTA for the majority of individuals in BCMC. Improvements to data quality and data processing are being made continuously to provide a more accurate and complete analysis of the PSA. In addition to regular reviews of the outcome measures, testing the predictive validity of the PSA for the local population is necessary to determine to what degree the PSA provides accurate predictions overall and if this accuracy varies by race or gender.

This study presented the Failure to Appear (FTA) Rate, the New Criminal Activity (NCA) Rate, and the New Violent Criminal Activity (NVCA) Rate for cases filed, assessed, closed, and those that had exposure in either Bernalillo County Metropolitan Court (BCMC) or Second Judicial District Court (SJDC) between July 1, 2017 and March 31, 2020. Also presented were correlations between risk factors and the outcome measures, details on new criminal activity, outcomes by race and gender, and details on predictive validity. We reviewed overall predictive validity as well as predictive validity by race and gender and judicial adherence to conditions.

The outcome measures for the PSA were calculated for the 10,289 cases in the study by recommendation category. As the FTA and NCA scores increased, so did the percent of failures. The overall FTA Rate was 22.8%. The FTA Rate generally increased as the FTA score from the PSA increased. The NCA Rate also increased as the NCA score increased, with an overall NCA Rate of 19.0%. The NVCA Rate was highest among those cases with an NVCA flag, and the overall NVCA Rate was 4.7%. Of the NCAs that occurred, nearly half (47.6%) were misdemeanors and another quarter were 4th degree felonies (25.8%).

While there was variation for the FTA, NCA, and NVCA scores by race, the difference in scores by group was not statistically significant. There were also no statistically significant differences between males and females in the FTA. There was a difference that was statistically significant for both the NCA Rate and NVCA Rate for males and females and the effect sizes were large and medium/large, respectively.

Additional comparisons were completed to determine if there were variations by race or gender in the FTA, NCA, or NVCA rates when conditioned on the associated flag or score: that is, whether defendants in one demographic group have lower or higher failure rates than other defendants with the same score. Such differences, if they exist, would suggest that the PSA overestimates or underestimates risk for that demographic group relative to the rest of the population. For FTA, there were no statistically significant differences by race or gender. By race, the NCA differences were not statistically significant in any individual score category, although using an additional technique that combines results across score categories (the Neyman smooth test) we found weak evidence, with p-value 0.002, that NCA Rates for Native American defendants are lower than the rest of the population when conditioned on NCA score. As further data becomes available, follow-up studies should be done to see if this pattern persists. Similarly, gender differences in NCA Rates were not statistically significant for any specific score but based on the Neyman smooth test, male defendants have significantly higher NCA Rates (meeting our threshold p=0.001) when conditioned on NCA score. For NVCA, there are no significant racial differences; for gender, the NVCA rate is significantly higher for male defendants, with a large effect size, but only when the flag is absent. When the NCA flag is present, male and female defendants have similar NVCA rates.

The predictive validity of the PSA was assessed using Area Under the Curve (AUC) of the Receiver Operator Characteristics (ROC) estimates. The AUC is a measure generally used to assess the ability of a tool to correctly identify positive or negative subjects. In general, the AUC scores for the PSA in Bernalillo County are similar to those seen in other jurisdiction, though the NVCA scores are slightly lower likely due to the use herein of the NVCA flag in lieu of the raw score. The PSA provides fair to good categorization of individuals, both overall as well as by gender for the FTA, NCA, and NVCA. The differences in scores were not statistically significant. This indicates that the PSA provides a more accurate assessment than random selection alone and that there is not a significant difference between race categories or by gender in its effectiveness.

Visually, the odds ratio figures support the previous statistical tests that showed considerable overlap between categories for the FTA, NCA, and NVCA Rates by race as did the FTA Rates by gender. Additionally, the fitted intervals for NCAs for females were lower and for NVCAs for females without the flag were lower as well. This supports the overall statistical findings in this report.

Overall, the key findings of this report indicate that while differences occur by race, these differences, some of which are noteworthy, were not statistically significant. Differences by gender in scores occur for NCAs and for NVCAs for individuals without the NVCA flag. The predictive validity of the PSA is demonstrably comparative to other jurisdictions and scores are comparable across groups, indicating that the tool is useful regardless of the individual's race or gender. As the available data continues to grow, methods become more refined, and local policies change, the continued analysis of the PSA will serve as a measure of the continued usefulness of the PSA in Bernalillo County.

We offer several recommendations:

1. The PSA should be revalidated every two years as recommended by the National Association of Pretrial Services Agencies that the Administrative Office of the Courts (AOC) "should review its risk assessment routinely to verify its validity to the local pretrial individual population… While there is no hard and fast rule here, this Standard recommends that pretrial services agencies conduct validations at least every two to three years" (NAPSA, 2020). This should be done by jurisdiction to address potential differences in populations as well as changes in jurisdictions over time in systems and procedures. The PSA tool is transparent in how the scores are calculated, what data is collected, and how risk factors are weighted and combined, allowing for in-depth analysis.
2. Distribute regular reports on outcomes to stakeholders so the outcomes by category are available to provide a broad understanding of failure and success by category that is consistent across agencies.
3. Periodically review the DMF and consider revisions based on actual success rates. While the PSA's FTA and NCA scores are produced by a mathematical process, the DMF that translates those scores into recommended conditions of release is a matter of policy. The DMF varies from jurisdiction to jurisdiction, and multiple jurisdictions have amended it over time as they learned more about the success and failure rates of defendants in each score category. The rates of FTA, NCA, and NVCA outcomes, both in Kentucky and in this study, did not always correspond well to the original dataset used by the Arnold Foundation to develop the PSA. Specifically, in all NCA score categories, NCA rates in New Mexico are significantly lower than those in the Arnold dataset. As a result, if the current DMF is based on Arnold's original materials, it may be based implicitly on an overestimate of risk. We also found that the DMF does not always correspond to observed risk: for instance, categories I, J, M, and N are statistically indistinguishable, all with FTA and NCA Rates of about 25% and 20% respectively, but the current DMF assigns three different recommended CORs to them (PML 1, PML 2, and PML 3). As part of a regular review, New Mexico stakeholders may choose to increase or decrease some recommended CORs to bring them in line with observed rates.
4. Investigate current methods of court notifications and invest in interventions such as text message reminders and study their effectiveness. Most FTAs (86.2%) involved only one missed hearing, and the majority (56%) consisted of missing only a preliminary hearing. This suggests that the FTA Rate could be significantly improved through interventions that encourage court appearance, especially in the early stages of the process. While it is unclear why these individuals still fail to appear, despite the large number of them on pretrial supervision, it is recommended to determine the state of notifications currently occurring in that program. In addition, recent research (Cooke et al., 2020; Fishbane et al., 2020) has shown that a well-designed and well-timed series of text messages that automatically remind individuals of upcoming hearings, ask them to make a plan to appear, and

alert them that failing to appear could lead to a warrant for their arrest, can reduce FTA Rates by 26% and lead to a 32% decrease in open warrants. We recommend that BCMC, SJDC, and other jurisdictions in the State of New Mexico consider a program of making these reminders available to all individuals and study the extent to which individuals make use of them and succeed in appearing.

5. Consider using the PSA for misdemeanor cases as well as felony cases.

6. Continue the commitment to transparency and avoid proprietary risk assessment tools. Unlike some competing risk assessment tools, the PSA is transparent in the sense that everyone can understand how the scores are calculated: what data from criminal records goes into it, and how the risk factors in this data are weighted and combined. This transparency is vital in criminal justice, where defendants, victims, and other stakeholders should have a clear understanding of why particular conditions of release were recommended. Moreover, transparency allows a deeper validation of a tool, measuring not just its overall accuracy but understanding its internal logic, including how it weighs and combines various risk factors such as previous convictions or failures to appear. Again, quoting the NAPSA Standards on Pretrial Release (p. 32) "any pretrial risk assessment instrument must be... transparent about its risk factors and their weighting" and preferably be in the public domain (p. 29-30). If this internal logic were hidden behind a veil of intellectual property, we would be much more limited in our ability to audit and validate it, and policymakers would be less able to make informed decisions about whether and how to use it.

References

AdvancingPretrial.org. (2020). About the Public Safety Assessment (PSA). Retrieved from:
https://advancingpretrial.org/psa/about/.

Arnold Ventures. (2019). Public Safety Assessment FAQs (PSA 101). Retrieved from
https://craftmediabucket.s3.amazonaws.com/uploads/Public-Safety-Assessment-
101_190319_140124.pdf.

Cooke, B., Diop, B. Z., Fishbane, A., Hayes, J., Ouss, A., & Shah, A. (2018). Using Behavioral Science to
Improve Criminal Justice Outcomes: Preventing Failures to Appear in Court. Ideas42 and
University of Chicago Crime Lab. Retrieved from: https://www.ideas42.org/wp-
content/uploads/2018/03/Using-Behavioral-Science-to-Improve-Criminal-Justice-Outcomes.pdf.

DeMichele, M., Baumgartner, P., Wenger, M., Barrick, K., Comfort, M. & Misra, S. (2018). The Public
Safety Assessment: A Re-Validation and Assessment of Predictive Utility and Differential
Prediction by Race and Gender in Kentucky. Retrieved from:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3168452.

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

Ferguson, E., Schwartz, T., & Guerin, P. (2018). Bernalillo County Public Safety Assessment Preliminary
Review. Institute for Social Research, *Unpublished*.

Ferguson, E., De La Cerda, H., & Guerin, P. (2019). Bernalillo County Public Safety Assessment Review –
July 2017 to March 2019. Institute for Social Research. Retrieved from:
http://isr.unm.edu/reports/2019/bernalillo-county-public-safety-assessment-review-july-2017-
to-march-2019.pdf.

Ferguson, E., De La Cerda, H., & Guerin, P. (2020). Failure to Appear and New Criminal Activity: Outcome
Measures for Preventive Detention and Public Safety Assessments. Retrieved from:
https://isr.unm.edu/reports/2020/failure-to-appear-and-new-criminal-activity-outcome-
measures-for-preventive-detention-and-public-safety-assessments.pdf.

Fishbane, A., Ouss, A., & Shah, A. K. (2020). Behavioral Nudges Reduce Failure to Appear for Court.
Science Magazine. Science 10.1126/science.abb6591. Retrieved from:
https://science.sciencemag.org/content/370/6517/eabb6591.

Ghosh, A., & Bera, A. K. (2001). Neyman's Smooth Test and Its Applications in Econometrics.
Retrieved from: https://ssrn.com/abstract=272888.

Greiner, D. J., Stubenberg, M., & Halen, R. (2020a). Validation of the PSA in Harris County, TX. Access to
Justice Lab, Harvard Law School. Retrieved from: https://a2jlab.org/public-safety-assessment-
validation-studies/.

Greiner, D. J., Stubenberg, M., & Halen, R. (2020b). Validation of the PSA in McLean County, IL. Access to
Justice Lab, Harvard Law School. Retrieved from: https://a2jlab.org/public-safety-assessment-
validation-studies/.

Greiner, D. J., Stubenberg, M., & Halen, R. (2021). Validation of the PSA in Kane County, IL. Access to
Justice Lab, Harvard Law School. Retrieved from: https://a2jlab.org/public-safety-assessment-
validation-studies/.

Laura and John Arnold Foundation. (2016). Public Safety Assessment: Risk Factors and Formula.
Retrieved from: https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/PSA-Risk-Factors-
and-Formula.pdf.

National Association of Pretrial Services Agencies. (2020). Standards on Pretrial Release: Revised 2020.
Retrieved from: https://napsa.org/eweb/DynamicPage.aspx?Site=napsa&WebCode=standards.

N.M. Constitution, art. II, §13.

Pretrial Justice Institute. (2015). Pretrial Risk Assessment: Science Provides Guidance on Assessing
Defendants [Issue Brief]. Retrieved from:
https://www.ncsc.org/__data/assets/pdf_file/0016/1654/pretrial-risk-assessment-science-provides-guidance-on-assessing-defendants.ashx.pdf.

Schlesinger, T. (2005). Racial and Ethnic Disparity in Pretrial Criminal Processing. *Justice Quarterly, 22:2, 170-192*. doi: 10.1080/07418820500088929.

Steffensmeier, D. J., Ulmer, J. & Kramer, J. H. (1998). The Interaction of Race, Gender, and Age in
Criminal Sentencing: The Punishment of Being Young, Black, and Male. *Criminology, 36(4), 763-798*. doi: 10.1111/j.1745-9125.1998.tb01265.x.

U.S. Constitution. amend. V.