



THE UNIVERSITY OF
NEW MEXICO

Bernalillo County Public Safety Assessment Validation Study, Report in Brief

June 2021

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

Prepared by: Elise Ferguson, M.A., Helen De La Cerdá, B.A., Paul Guerin, Ph.D., and Christopher Moore, Ph.D.

In this brief: This study reports outcome measures associated with the PSA and assesses the ability of the PSA to accurately identify the likelihood of court appearance and new criminal activity.

Overall Pretrial Outcomes

Release Rate: 69.1%

Appearance Rate: 77.1%

Public Safety Rate: 81.0%

New Violent Criminal Activity Rate: 4.7%

Highlights

The sample consisted of 10,289 assessed and closed felony cases with exposure between July 2017 and March 2020.

The Appearance Rate is defined as the rate at which released individuals attend all scheduled court appearances and was 77.1%. The Failure to Appear Rate was 22.9%.

The Public Safety Rate is defined as the rate at which released individuals do not have a new offense and was 81.0%. The New Criminal Activity Rate was 19.0%.

The New Violent Criminal Activity Rate was 4.7%.

Differences in FTA, NCA, and NVCA rates by race and FTA by gender were not statistically significant.

The NCA rate for males was nearly 5% higher than for females. The NVCA rate was 2.2% higher for males than females.

(Continued on page 2)

Background

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. In most jurisdictions PSAs are completed for individuals facing misdemeanor or felony charges. In Bernalillo County, PSAs are completed for individuals with felony charges only.

This report in brief summarizes the full report that is a validation of the PSA and was preceded by a number of reports that reported on the implementation and outcomes associated with the PSA. The PSA utilizes nine factors to predict failure for three pretrial outcome measures: New Criminal Activity (NCA), New Violent Criminal Activity (NVCA), and Failure to Appear (FTA). The PSA is scored by reviewing an individual's criminal history, current cases, and age to create an FTA and NCA score as well as a flag for NVCA. Bernalillo County, like other jurisdictions that utilize the PSA, developed a Decision Making Framework (DMF). The DMF plots the individual's NCA score against the individual's FTA score and provides a recommended release category for the individual.

The PSA was validated across scores, racial/ethnic groups, by gender, outcome measures of Failure to Appear (FTA), New Criminal Activity (NCA), and New Violent Criminal Activity (NVCA), and finally, the Adherence Rate. The adherence rate was compared to the outcome measures. The Appearance Rate and Public Safety Rate are considered the inverse of the FTA and NCA Rates. The release rate was also calculated for the

outcome measures sample. The rates are defined as follows:

- Appearance Rate – the rate at which individuals do not receive a warrant for failure to appear at a hearing
- Public Safety Rate – the rate at which individuals are not charged with a new offense during the pretrial stage
- Release Rate – the rate at which individuals are released during the pretrial stage
- Adherence Rate – the degree to which conditions of release (COR) correspond with the PSA recommendation category

Population Demographics

Most cases had a male defendant (74.5%). In 50.4% of cases, the defendant was Hispanic. The percentage of White defendants was 32.4%, followed by Native American and African American, at 7.1% and 6.9% respectively. The remaining 3.1% of individuals for cases in the sample were unknown, multi-racial, or another race category (see Table 1).

Table 1. 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%
African American	713	6.9%
Other	321	3.1%

PSA Outcome Measures

Any warrant issued for FTA at a scheduled court appearance was considered a valid warrant and applied to the FTA Rate. The FTA Rate was 22.9% and an Appearance Rate of 77.1%. As the recommendation category became more restrictive, the FTA Rate increased, and the Appearance Rate decreased.

(Continued from page 1)

For most NCAs, the charge level was lower (47.5%) or the same (39.3%) as the assessed case.

The Area Under the Curve (AUC), a general measure of the accuracy of predictive tools, did not vary significantly by race or gender.

Judicial adherence to PSA recommendations was 76.5%.

FTA, NCA, and NVCA rates were lowest when conditions of release were less restrictive or when they adhered to the PSA recommendation.

There were no differences in FTA rates by race or by gender.

There were no statistically significant differences in NCA by race.

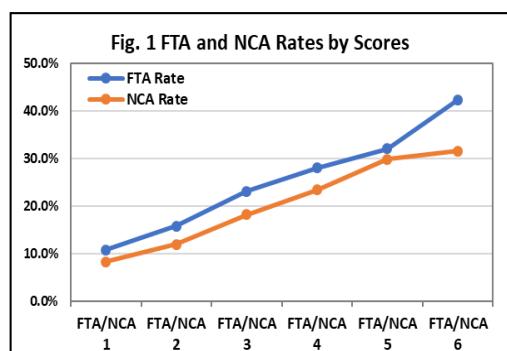
The NCA rates for Native American were low and did not meet the threshold for statistical significance, the p-value (0.002) approaches significance and may be important to review in future.

Differences in the NVCA rate by the NVCA flag by race were not statistically significant.

To a statistically significant degree men had higher NCA rates than women.

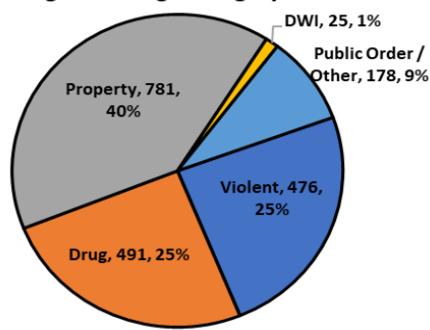
Differences in the NVCA rate by the NVCA flag by gender were statistically significant with a large effect size.

The NCA Rate for all cases was 19.0%. The Public Safety Rate ranged from 91.7% for cases with an NCA 1 to 68.4% for cases with an NCA 6. Similar to the FTA Rate, the NCA rate increased, and the Public Safety Rate decreased as the recommendation category became more restrictive.



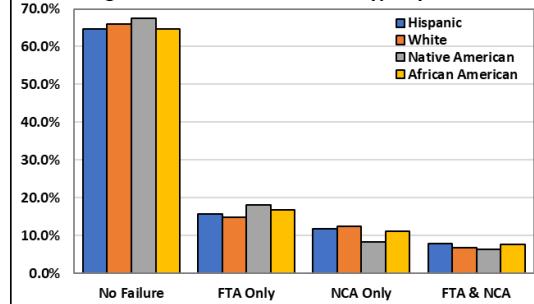
Sixteen percent of the sample cases (1,650) were assigned an NVCA flag on the PSA. The NVCA Rate for all cases was 4.7% with 95.3% of individuals not having a new violent offense. Almost half (47.5% or 926) of the cases with an NCA had a charge level lower than the case for which they were assessed. This is an important finding and shows that while 19.0% of individuals had an NCA during the pretrial period, it was most often a lower-level charge or an equivalent level charge (39.3% or 766). Figure 2 reports the highest charge category for NCA cases.

Fig. 2 Highest Charge Category of NCA Case



Figures 3 and 4 display the overall failure or success by race and gender. Individuals who succeeded—that is had no FTA or NCA during the pretrial phase—did so at a rate between 64.7% for African Americans up to 67.5% for Native Americans. The range for individuals with only an FTA was 3.1% and for those with only an NCA was 4.1%. Between 6.3% of Native American individuals and 7.8% of Hispanic individuals had both an FTA and an NCA.

Fig. 3 Overall Success or Failure Type by Race



METHODOLOGY

The sample was constructed using electronic court data from the Bernalillo County Metropolitan Court and the Second Judicial District Court. The electronic data included 29,645 cases filed between July 1, 2017 and March 31, 2020. Cases were included for analysis if they met the following criteria:

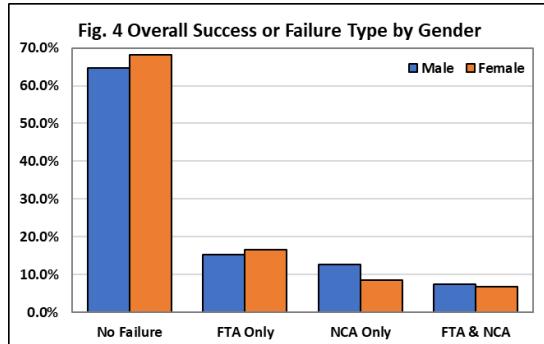
1. A PSA was completed
2. The defendant was in custody for the Felony First Appearance (FFA) or Felony Arraignment (FA)
3. The case was closed by March 31, 2020
4. The defendant was released during the pretrial period

There were 10,289 cases fit the study criteria and were included for analysis.

These data were used to assess the predictive

validity, uniform validity, and equitable validity using Area Under the Curve (AUC) Receiver Operating Characteristic (ROC), Chi-Square Test, Pearson's Correlation, the Neyman Smooth Test, 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.



Female individuals succeeded during the pretrial phase 3.5% more of the time than males (68.1% compared to 64.6%). While female individuals had a slightly higher percent of those with only an FTA by 1.3%, male individuals had had a higher rate of those with only an NCA by 4.1%. The rate of those with both an FTA and an NCA was nearly the similar for males and females at 7.5% and 6.8% respectively.

Overall, there are no significant differences in FTA, NCA, or NVCA rates by race, or in FTA rates by gender. NCA rates for Native American defendants are lower than others with the same NCA score, but due to the small sample size this difference may not be statistically significant, and this should be reexamined as more data becomes available. The NCA and NVCA rates were higher for males than females and this difference was statistically significant (by 4.8% and 2.2% respectively).

Predictive Validity of the PSA

This section describes the tool's predictive validity, uniform validity, and equitable validity. The Area Under the Curve of the (AUC) Receiver Operator Characteristics curve is a measure of a tool's accuracy that is often used in risk assessments or medical diagnoses. It is a measure of how the sensitivity and specificity varies with the threshold we use to label a subject positive or negative. The AUC score ranges from 0.0 to 1.0, and 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.

The AUC ROC estimates were calculated for FTA, NCA, and NVCA. For the FTA, the AUC score was 0.641. The AUC score was slightly lower for NCA (0.638) and NVCA (0.572). The AUC scores for FTA and NCA were considered good and the predictive validity of the PSA for NVCA was fair. These scores indicate that for all three outcomes, the PSA provides more accurate estimates for pretrial success than random chance alone.

The AUC scores indicate 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. Table 2 displays the AUC scores for each outcome measure by race and gender.

There were no significant differences in AUC scores by race or gender. There were also no statistically significant differences in the FTA, NCA or NVCA rates by race or gender. The only significant findings were the differences in the NCA – which did not vary significantly in any individual score category – and NVCA rates by gender, both of which were significant differences with a notable effect size.

Adherence

Adherence describes the degree to which the ordered conditions of release (COR) as recommended by the PSA correspond to the COR ordered by the judge. Only cases that had COR ordered were included in the adherence review. Cases for which there was a preventive detention motion were excluded. There were 1,167 cases excluded from the 10,289, leaving 9,122 cases for adherence review.

Overall, the conditions set 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. Of the cases where the individual had Detain/Max conditions recommended, the judge ordered additional conditions 22.5% of the time and less restrictive conditions 3.9% of the time.

The FTA rates were highest when the COR were more restrictive than the recommendation (27.9%), and lowest when the COR were less restrictive (17.9%).

The NCA rates were also highest for those with more restrictive conditions than the recommendation (21.6%) and lowest for those with less restrictive conditions (15.3%).

Finally, the NVCA rate for those with more restrictive conditions was 4.9%, 4.3% when the conditions adhered, and 3.4% when the conditions were less restrictive.

At the lower FTA and NCA score levels, success rates were roughly equal regardless of adherence. At higher score levels, more restrictive COR were associated with higher rates of FTA and NCA, while adherence or less restrictive COR were associated with lower rates of FTA and NCA. There was no significant association between adherence and NVCA.

Table 2. AUC Scores by Outcome Measure, Race & Gender

Outcome Measure	AUC Score (N=10,289)	Hispanic (N=5,188)	White (N=3,334)	Native American (N=733)	African American (N=713)	Male (N=7,670)	Female (N=2,619)
FTA	0.641	0.642	0.628	0.614	0.683	0.643	0.638
NCA	0.638	0.63	0.636	0.645	0.661	0.639	0.623
NVCA	0.572	0.558	0.576	0.619	0.543	0.569	0.568

Discussion

This is the first validation study of the PSA in Bernalillo County since the implementation in 2017. Validation studies are vital, and more research needs to occur that focuses on reporting 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. 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.

The predictive validity of the PSA was assessed using AUC ROC estimates. The AUC is a measure generally used to assess the ability of a tool to correctly distinguish positive or negative subjects. The PSA provides fair to good categorization of individuals, both overall as well as by race and by gender for FTA, NCA, and NVCA. The differences in scores were not statistically significant, which indicates the PSA provides a more accurate assessment than random selection alone and shows there is not a statistically significant difference between race categories or by gender in its effectiveness. 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.

The AUC scores for the PSA in Bernalillo County were similar to validity findings in other jurisdictions. The PSA in other jurisdictions generally provided fair to good categorization in Kentucky (DeMichele, Baumgartner, Wenger, Barrick, Comfort & Misra, 2018), McLean County, Illinois (Greiner, Stubenberg, & Halen, 2020b); Harris County, TX (Greiner, Stubenberg, & Halen, 2020a); and Kane County, Illinois (Greiner, Stubenberg, & Halen, 2021).

While the findings in Bernalillo County were similar to findings in other jurisdictions, this study contributes to a more detailed review of the NCA charges include the level and category, a comparison of the level with the source charge level, and adherence to conditions of release.

The outcome measures for the PSA were calculated by recommendation category. As the FTA and NCA scores increased, so did the percent of failures. The overall FTA Rate was 22.8% and as FTA score from the PSA increased so did the FTA Rate. The NCA Rate overall was 19.0% and increased as the NCA score increased. Additionally, NCA Rates in Bernalillo County by NCA score categories were smaller than those reported by Arnold Ventures using their original dataset. 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%).

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. The findings of this report should be used to inform and drive additional policy decisions. Additionally, it is important to look at local pretrial, jail, and court processes. 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.

RECOMMENDATIONS

1. The PSA should be revalidated every two years as recommended by the National Association of Pretrial Services Agencies.
2. Information on how the PSA is performing should be distributed on a regular basis to stakeholders and the general public. This includes the adherence rate and the outcome measures.
3. In addition to revalidation, we suggest an in-depth study of the DMF and potential revisions of the DMF based on those outcomes. As part of this in-depth study, 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.
5. The criminal justice system should consider using the PSA for misdemeanor cases.
6. Continue the commitment to transparency and avoid proprietary risk assessment tools.

The UNM - Institute for Social Research (ISR) is a leading provider of program evaluations and policy research in New Mexico. ISR staff members and faculty affiliates also have expertise in the fields of criminal justice, education, economics, substance abuse treatment programs, poverty and homelessness, domestic violence, employee workloads and staffing levels. For more information on the ISR, please visit <http://isr.unm.edu/> or call (505) 277-4257.