

UNIVERSITY OF NORTHERN COLORADO

Colorado Pretrial Assessment Tool (CPAT) Validation Final Report Summary

Victoria A. Terranova, PhD¹

Kyle C. Ward, PhD

Department of Criminology and Criminal Justice
University of Northern Colorado

The CPAT Validation Study began in January of 2018 with the goal of empirically validating the current version of the CPAT, gathering insight from pretrial practitioners and stakeholders throughout Colorado about their perceptions of the current tool and their suggestions to improve it, and creating, piloting, and validating a revised version of the tool (The CPAT-Revised (CPAT-R)). The study included data from seven counties (Boulder, Denver, Larimer, Garfield, Mesa, Pueblo, and Weld), as well as survey responses from pretrial practitioners across the state. Additional data sources include Colorado Judicial Branch Courts, Denver Municipal Court Records system, and the Colorado Bureau of Investigation. The key findings from the third phase of the report are outlined below. For more in depth description and analysis, please refer to the final report (<https://www.unco.edu/hss/criminology-criminal-justice/pdf/final-report-cpatr-project.pdf>).

The findings from the third phase of the study supported a recommendation of the CPAT-R. The recommended CPAT-R was produced from modifying the piloted CPAT-R for equity in prediction. The recommended CPAT-R was estimated to be the most accurate, tested and balanced for equity in predictive performance across defendant sub-groups including race/ethnicity, sex, and residential status. The differences in predictive performance across sub-groups were not statistically significant. This means that the performance in the recommended CPAT-R does not meaningfully differ across sub-groups. The descriptive differences in the rate of high-risk errors (false positive rate) were relatively reduced to less than a 10% difference across defendant race/ethnicity and sex. The difference in high risk error across defendant residential status was reduced relatively to a 21% difference.

Phase 3: Construction, Pilot, and Validation of the CPAT-R

The third phase of this project included the construction, pilot, and subsequent assessment of a modified version of the CPAT, the CPAT-Revised (CPAT-R). The CPAT-R was piloted over 3-months (November 13, 2018 - February 13, 2019) in the seven participating counties. The CPAT-R interview was administered in addition to the conventional CPAT interview to a randomly assigned sub-group of pretrial defendants and subsequently scored by Pretrial Services Officers.

¹ Correspondence can be sent to victoria.terranova@unco.edu or kyle.ward@unco.edu

The recommended CPAT-R was constructed to assign risk of the primary pretrial outcome, new arrest and/or Failure to Appear (FTA). This is defined as an arrest for an offense that occurred during the pretrial release time period. This includes arrests for misdemeanor and felony-level charges, as well as new charges in the Denver Municipal Court Records system. This outcome is measured using court records to define the date of arrest as well as offense and to identify new arrests during the pretrial release time period. Offenses excluded from new arrest include traffic offenses and arrests for FTA. FTA is measured by the court event records. It is defined as a failure to appear to court for a scheduled court hearing/date during the pretrial release time period. The new arrest and/or FTA outcome is defined either as a new arrest, an FTA or both occurring during the pretrial release time period.

The following is a breakdown of the primary pretrial outcomes of the full pilot sample:

- Total sample size of released pretrial defendants: 3,757
- New Arrest or FTA - at all: 35.59%²
- New Arrest – at all: 19.70%
- FTA – at all: 22.65%

The following details the primary findings from phase three and the final recommended tool that are informed by the findings from all three phases of the study.

Predictive Performance

Predictive performance was assessed for the validity and accuracy of the CPAT-R, CPAT and CPATR-SV. Predictive performance is the ability of the tool to correctly predict whether a defendant will get arrested or FTA while released on bond. The Receiver Operating Characteristic (ROC) and corresponding Area Under the Curve (AUC) score were estimated to assess predictive performance. This represents the ratio of times that the risk score is and is not consistent with its anticipated outcome. Validated pretrial risk assessments AUC scores range between .50 and .75.³ Predictive performance was also assessed according to the tool's calibration. This represents the likelihood that a risk score is consistent with the predicted outcome.

- The piloted CPAT-R (AUC = .65), CPAT (AUC = .60) and CPATR-SV (AUC = .62) validated, demonstrating each accurately assign risk of new arrest and/or FTA. The differences in accuracy were statistically significant. This finding supports that the

² New arrest or FTA – at all, $n = 1,337$; New arrest – at all, $n = 740$; FTA - at all, $n = 851$

³ Validation scores are represented by the Area Under the Curve (AUC). The AUC, which is estimated using the Receiver Operating Characteristic (ROC) is used to infer the probability that the risk assignment will assign a score consistent with the pretrial outcome. Validation scores can indicate a degree of predictive performance. Pretrial assessment tools with fair performance have AUC scores ranging .55-.63, good .64-.70 and excellent is .71 or higher. See Desmarais, S.L. & Singh, J.P. (2013.). *Risk assessment instruments validated implemented in correctional settings in the United States*.

piloted CPAT-R was the most accurate tool and best candidate for the subsequent bias analysis and balancing modifications for predictive equity.

- After balancing for bias in performance the recommended CPAT-R validated (AUC = .66) for the primary outcome, new arrest and/or FTA. It also validated for the other primary outcomes of new arrest (AUC = .57) and FTA (AUC = .69).
- The recommended CPAT-R was estimated to have the greatest predictive performance compared to the CPAT and CPATR-SV. The CPATR-SV is a version of the CPAT-R that does not require an interview. It contains only the risk factors that can be confirmed with official criminal history or court records. The recommended CPAT-R was validated in this study and is estimated to perform with comparable accuracy to other well-validated pretrial risk assessment tools.⁴

Bias Analysis

The bias analysis used self-reported demographic information to examine if the tool performed differently in relation to a pretrial defendant's race/ethnicity, sex, and residential status.⁵ The piloted CPAT-R was identified as the best candidate for the bias analysis because it was estimated to be the most accurate compared to the CPAT and CPATR-SV. The pilot CPAT-R was assessed for meaningful differences in predictive performance across the above identified sub-groups and the relative descriptive differences in the error rates across sub-groups. Modifications to the piloted CPAT-R were made to reduce any of these differences and tested across sub-groups on a different sample of pretrial defendants.

- The piloted CPAT-R was assessed for predictive performance across race/ethnicity, sex, and residential status. As a result, the piloted CPAT-R was modified to reduce any differences in the tool's performance across sub-groups. These modifications include removing two risk factors: 1) prior violent arrest and 2) the self-reported time living at current residence. Prior violent arrest was defined as a prior offense that a victim is harmed by or threatened by violence. Time living at current residence was defined as living at least one year at the defendant's current residence.
- The predictive performance of the CPAT-R across sub-groups was estimated using the ROC and corresponding AUC scores. No statistically significant differences were identified across the AUC scores indicating that the differences are not meaningful or likely to be replicated in the population of pretrial defendants.
- The descriptive differences in the assessment errors were estimated across sub-groups. The false positive rate was the highest priority assessment error, which represents the proportion of times that those assessed as high risk did not have a high-risk outcome. The relative differences in the recommended CPAT-R across race/ethnicity were reduced to

⁴ A local validation of the Public Service Assessment (PSA) was recently conducted estimating an AUC score of .65 (see DeMichele et al, 2018.). A well-established pretrial risk assessment tool, the ORAS-PAT, has also validated with an estimated AUC of .65 (see Latessa et al, 2010).

⁵ Race/ethnicity – white, black, Hispanic, sex – male, female, residential status – homeless, housed

.06 and gender were reduced to .05. The relative differences across residential status were slightly reduced to .21.

- The relationship between sub-group membership and the performance of the CPAT-R was also assessed. The interaction between membership and performance was estimated indicating whether the recommended CPAT-R's performance was influenced by sub-group membership. The interaction terms for this analysis were not statistically significant indicating that there were no systematic interactions between sub-group membership and the recommended CPAT-R's performance.
- Assessing bias in the performance of the CPAT and CPATR-SV, no statistically significant differences were identified across the AUC scores, indicating that there was no meaningful difference in the accuracy of either tool across sub-groups. Differences in the assessment errors across sub-groups were also estimated. These descriptive differences in the false positive rates for CPAT across sub-groups were estimated to be across: race/ethnicity- .11, sex- .07, and residential status- .28. The differences in the false positive rates for the CPATR-SV across sub-groups were estimated to be across: race/ethnicity- .09, sex- .01, and residential status- .08.

Reliability and Features

The reliability of self-reported information provided by the pretrial defendant was assessed by comparing this information with what is recorded in one's criminal history record. The level of agreement was analyzed for all risk factors that could be confirmed with official data including: prior arrest, arrest in the last year, prior FTA, FTA in the last year, pending charge at arrest, and active warrant at time of arrest.

Features of the risk factor definitions were also tested including self-reported current or past self-reported problem with alcohol or drugs. The predictive performance of the recommended CPAT-R was estimated without the past-only feature using the ROC and corresponding AUC score.

- All risk factors systematically agreed across self-report and official records at a low or moderate level except for prior arrest. Of the pretrial defendants that were interviewed for the CPAT-R, more self-reported having two or more prior arrests (95%) than what their criminal history indicated (73%).
- The recommended CPAT-R with current-only alcohol or drug risk item was validated (CV-AUC = .66, BC CI: .62-.68). The past-only feature of this risk factor was removed as a feature of this risk factor's definition from the implementation protocol.

Recommendations

The study's final recommendations are supported by the study findings from all 3 phases. The third phase recommendation pertain to the revised tool's construction and implementation. The recommended CPAT-R that was assessed for predictive performance and equity in performance across sub-groups. Estimates support that the recommended CPAT-R is the best performing and balanced tool across sub-groups.

Recommended CPAT-R

Risk Factor	Score	Definition
Employment/education	0/2	Self-reported employment or current student at the time of arrest. (0 = yes, 2 = no).
Current problems with alcohol or drugs	0/1	Self-reported current problems with alcohol and/or drugs (0 = no, 1 = yes).
Prior Arrests	0/3	Prior arrests confirmed with criminal history records (0 = 1 or less, 3 = 2 or more).
Arrest in the last year	0/3	Arrest within the last year confirmed with criminal history records (0 = none, 3 = 1 or more).
Age at first arrest	0/1	Defendant age at first arrest confirmed with criminal history (0 = 21 years old or older, 1 = 20 years old or younger).
Prior FTA	0/3	Prior FTA confirmed with court history records (0 = none, 3 = 1 or more).
FTA in the last year	0/3	FTA within the last year confirmed with court history records (0 = none, 3 = 1 or more).
Pending charge at arrest	0/1/2	Pending charge at arrest (0 = none, 1 = misdemeanor charge(s) only, 2 = at least 1 felony charge).
Active warrant	0/2	Active warrant at arrest (0 = no, 2 = yes).

a) Range 0 – 20

b) Category 1, score 0-7; Category 2, score 8-11; Category 3, score 12-14; Category 4, score 15-20

Category ^a	Rate of success – new arrest and/or FTA ^b	Odds of success – new arrest and/or FTA ^c	Rate of success – new arrest ^d	Odds of success – new arrest ^e	Rate of success – FTA ^f	Odds of success – FTA ^g
1	89%	4.60 greater	97%	8.48 greater	91%	10.01 greater
2	59%	1.44 greater	82%	4.58 greater	71%	2.36 greater
3	55%	1.20 greater	78%	3.50 greater	67%	2.02 greater
4	54%	1.18 greater	74%	3.24 greater	66%	1.97 greater

a) The percentage breakdown of the sample who fell in each category: 1) 41.73%, 2) 24.40%, 3) 17.17%, 4) 15.71% $n = 1,858$

b) The proportion of those in the risk category who are not arrested or FTA during the pretrial release period.

c) The odds of no new arrest and/or FTA vs. a new arrest and/or FTA occurring (e.g. Category 1 defendants have a 4.60 greater odds of having no new arrest and/or FTA while release on bond compared to having a new arrest and/or FTA).

d) The proportion of those in the risk category who are not arrested during the pretrial release period.

e) The odds of no new arrest vs. a new arrest occurring (e.g. Category 1 defendants have 8.48 greater odds of having no new arrest while released on bond compared to having a new arrest.)

f) The proportion of those in the risk category who do not FTA during the pretrial release period.

g) The odds of no FTA vs. an FTA occurring (e.g. Category 1 defendants have a 10.01 greater odds of having no FTA while released on bond compared to having an FTA occurring).

- *Accuracy and Balance:* The recommended CPAT-R should be evaluated for predictive performance and equity in performance on-going. It should also include evaluation of the fidelity to the tool's implementation protocol.
- *Appropriate and Standardized Use:* The recommended CPAT-R should be used to inform pretrial release decision-making **in conjunction** with professional discretion. This tool is constructed to assign a total score that represents the risk of the primary pretrial outcome, new arrest and/or FTA. It should be used to inform decisions related to these pretrial release outcomes. Training and education of those that both administer to tool and use it to inform decision-making is recommended to enhance the standardized use and interpretation of the tool that is consistent with its construction and protocol.