Colorado Pretrial Assessment Tool Validation Project
Final Report

Victoria A. Terranova, PhD

Kyle C. Ward, PhD

July 1, 2020

Department of Criminology and Criminal Justice

University of Northern Colorado

The project has been funded by The State of Colorado Office of State Planning and Budgeting and the University of Northern Colorado

Acknowledgements

This study was employed through a collaboration with Colorado's Pretrial Executive Network (PEN). The PEN group is a collective of pretrial executives throughout Colorado who share pretrial-related information, research, and tactics. It was through this group that this study was developed. We would like to personally thank Doug Erler, Steve Chin, Gary Darling, Greg Mauro, Monica Rotner, Elizabeth Stewart, and Brian Gomez for all of their support throughout the study. We would also like to thank the State of Colorado's Office of State Planning and Budgeting and the University of Northern Colorado for providing the financial support and assistance to complete the study. In addition, and in no particular order, we would like to thank the following individuals who work throughout Colorado's criminal justice system: Kim English, Peg Flick, Aubree Cote, Alison Birchard, Nathan Thorn, Chad Dilworth, Sri Komakula, Kamie Cooley, Bobbi Vigil, Judith Bryant, Linette Schweizer, Allison Tapia, Stacey Conrardy, Shawn MacNamara, Lindsay Kelly, Derik Stalls, and Stan Hilkey. Your assistance with data collection, observation, and providing feedback were paramount to the successful completion of this study. Included in this list are also those at the Colorado Judicial Branch and Colorado Bureau of Investigation that assisted providing data. We are incredibly fortunate for the support of our team at the University of Northern Colorado. We would like to thank our dedicated graduate students who were instrumental in the data collection and data entry efforts at multiple points throughout this study: Anthony Azari, Jessie Slepicka, Samantha Lytle, and Annaliese Engler. Finally, we would like to thank all of the pretrial officers in the counties of Boulder, Denver, Garfield, Larimer, Pueblo, Mesa and Weld who provided their time and effort implementing the pilot of the CPAT-R.

Table of Contents

Tables

Executive Summary

The Colorado Pretrial Assessment Tool (CPAT) is an empirically-based pretrial risk assessment tool developed to inform bond decisions in Colorado. The tool measures a defendant's risk of failure to appear (FTA) or re-arrest while released pretrial. The overall goal of this study was to examine and validate the current CPAT, and based on feedback from multiple criminal justice stakeholders, construct and pilot a revised version of this tool, the CPAT-Revised (CPAT-R). The study was completed between January 2018 and June 2020 and included three phases. This report will detail the methodology and results of all three phases of this study. Below is a summary of the results of each of the three phases.

Phase One

Phase one includes a retroactive validation of the CPAT covering 3,386 cases from 2015 and 2016. Data from this study were provided by each of the seven participating counties (Boulder, Denver, Garfield, Larimer, Mesa, Pueblo, and Weld), the Colorado Judicial Branch, Denver Municipal Court Records system, and the Colorado Bureau of Investigation. In addition, an online survey was administrated to criminal justice stakeholders across the state. A total of 382 respondents, representing 27 counties, completed the survey. These stakeholders consisted of representatives from pretrial services, judges, public defenders, and district attorneys. The following are key findings that will be discussed in greater detail throughout this document.

- The CPAT validated showing to assign the correct risk of new arrest and failure to appear.
    - Validated pretrial risk assessments receive a validation score between .50 and .75. The CPAT's validation score was .58 for new arrest or FTA.[1]
- Further examination of risk category designation and a modified tool could improve the predictive performance of the CPAT.
    - Preliminary analysis indicates that modifications of the tool could increase the validation score to above .60.
    - Features of the tool that could possibly be modified include: risk categories, risk items, and specification of risk outcomes.
- The method of weighting and scoring the risk factors of the current CPAT does not interfere with the predictive performance of the CPAT.

[1] 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 setttings in the United States.* 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).

- Results from a survey of key stakeholders identified implementation issues with the current CPAT that include: lack of account of prior failure to appears, lack of a substance abuse question, reliability of self-report items, overload of Category 2, lack of account of current charge severity, and buy-in among different stakeholders.

The following is a breakdown of the primary pretrial outcomes of the 2015 – 2016 retroactive validation sample:

- Total sample size: 3,386
- New Arrest or FTA - at all: 43%[2]
- FTA – at all: 30%
- New Arrest – at all: 29%

Risk category breakdown of the CPAT risk category assignment of the 2015 – 2016 retroactive validation sample:

- Category 1: 11.06%
- Category 2: 36.91%
- Category 3: 29.26%
- Category 4: 22.78%

Phase Two

The second phase of this project included focus groups and observations of the seven participating counties including: Boulder, Denver, Garfield, Larimer, Mesa, Pueblo and Weld. Participants for the focus groups were recruited from a preliminary survey disseminated in phase one and through recruitment emails from pretrial directors. A total of 14 focus groups were held between May 10, 2018 and June 30, 2018. There were two categories of focus group participants: 1) pretrial officers, and 2) criminal justice stakeholders that used risk assessment tools to inform decision-making and bail bond arguments (i.e., judges, prosecutors, public defenders, and other criminal justice administrators). In total, there were 41 pretrial officers and 68 stakeholders that participated in the focus groups.

A semi-structured interview guide was used with questions that pertained to how the participants perceived the CPAT in its current state, as well as the utility of the tool for their job duties. Other questions were asked about perceptions of buy-in of the tool, training, and potential modifications to the tool. Thematic content analysis was employed on transcriptions of the focus groups and five preliminary themes were identified from the analysis.

[2] New arrest or FTA – at all, n = 1,354 ; FTA - at all, n = 1,015; New arrest – at all, n = 903

Emergent themes:

- *Role of pretrial risk assessment*

  Pretrial officers and stakeholders perceived the risk assessment tool to have value in the pretrial process. Since these participants have different roles, the value of risk assessment differs for each of their decision-making processes.

- *How the tool is used*

  Primarily, pretrial risk assessment is perceived as most useful for informing decisions about pretrial release. Although pretrial risk assessment was perceived as useful, additional factors beyond the tool were also referenced as important considerations of the release decision.

- *Consideration of other factors, independent of the risk assessment*

  A variety of considerations were cited by participants as relevant to pretrial decision-making that were independent of pretrial risk assessment. Some of these factors included: history of failure to appear to a scheduled court date, arrest type, and factors representative of a pretrial defendant's community stability.

- *Pretrial supervision decisions*

  It was reported by pretrial decision-makers that bond release and supervision condition decisions are often made at the same time. Pretrial risk assessment scores were perceived to be useful for informing both of these steps in the pretrial process.

- *Training and Education*

  Both types of focus group participants, pretrial officers and the pretrial stakeholders perceived on-going training and education about the tool to be important. Pretrial officers reported receiving a greater amount of training and education than pretrial stakeholders. Pretrial stakeholders reported a need for enhancing the current amount of training and education they receive on pretrial risk assessment.

Phase Three

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). This phase was informed by the preliminary findings from phase one that indicated the potential for improvement to the CPAT and findings from phase two about the tool's construction and implementation. 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 sub-group of pretrial defendants and subsequently scored by pretrial officers.

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%[3]
- New Arrest – at all: 19.70%
- FTA – at all: 22.65%

This information was then used to assess the predictive performance of the CPAT-R in comparison to the CPAT in its current state. A version of the CPAT-R that would not require a pretrial assessment interview, the CPATR-Screening Version (CPATR-SV) was also assessed. The total pilot study sample included 5,263 interviewed pretrial defendants of which, 3,757 were released pretrial. Predictive performance is assessed by estimating the accuracy of the tool's risk assignment to the pretrial outcome that occurs. The best performing tool is then assessed for accuracy equity and predictive parity across sub-groups.[4] Finally, the features of the risk factor definitions of the tool identified as the best performing and equitable across sub-groups were assessed.

These findings are used to inform inferences about the best performing and most equitable tool that can be used for pretrial release decisions in the state of Colorado. The following details the primary findings from phase three and the final recommended tool informed by the findings from all three phases of the study.

- The piloted CPAT-R, CPAT and CPATR-SV validated showing to accurately assign risk of new arrest and FTA. The piloted CPAT-R was estimated to perform meaningfully better than the other tools, with an AUC score of .65.[5]
- The piloted CPAT-R was assessed for accuracy equity, and predictive parity across race/ethnicity, gender, and residential status sub-groups. This assessment informed a modification to the piloted CPAT-R that resulted in removing two of the risk factors: 1) prior violent arrest and 2) the self-reported time living at current residence.
- Assessment of the features of the risk factor definitions informed further modification to the tool's risk factor definitions including removing self-reported prior alcohol or drug problem as scored response. Current problems with alcohol or drugs remains a validated item in the CPAT-R.

[3] New arrest or FTA – at all, n = 1,337; New arrest – at all, n = 740; FTA - at all, n = 851

[4] Accuracy equity refers to the predictive performance of an assessment tool across sub-groups. Predictive parity refers to differences in assessment error across sub-groups.

[5] AUC = .65 (BC-CI: .62-.68). Relative to other risk assessment tools, an AUC score of .65 would indicate good predictive performance. See Desmarais, S. L., and Singh, J P. (2013). *Risk assessment instruments validated and implemented in correctional settings in the United States.* Lexington, Kentucky: Council of State Governments. 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).

- The recommended CPAT-R was re-validated on a test sub-sample of the primary pretrial outcomes: new arrest and/or FTA, new arrest, and FTA. The test sample's assessment errors were comparable indicating accuracy equity across race/ethnicity and gender, and minimal bias across sub-groups. Differences in predictive parity across residential status were reduced but were still greater than the other sub-group categories.

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

Risk Category	Rate of success – new arrest[a]	Odds of success – new arrest[b]	Rate of success – FTA[c]	Odds of success – FTA[d]
1	97%	8.48 greater	91%	10.01 greater
2	82%	4.58 greater	71%	2.36 greater
3	78%	3.50 greater	67%	2.02 greater
4	74%	3.24 greater	66%	1.97 greater

a) The proportion of those in the risk category who are not arrested during the pretrial release period. *n* = 1,858
b) The odds of no new arrest vs. a new arrest occurring (e.g. Category 1 defendants have an 8.48 greater odds of no new arrest compared to a new arrest.)
c) The proportion of those in the risk category who do not FTA during the pretrial release period.
d) The odds of no FTA vs. an FTA occurring (e.g. Category 1 defendants have a 10.01 greater odds of no FTA compared to an FTA occurring).

Project Background

The CPAT was created in 2012 as part of the Colorado Improving Supervised Pretrial Released (CISPR) Project.[6] The development of the CPAT entailed a sample of 2,000 defendants from 10 Colorado counties (Adams, Arapahoe, Boulder, Denver, Douglas, El Paso, Jefferson, Larimer, Mesa, and Weld). Through the analysis of 177 collected variables, the 12-item CPAT was created and adopted by additional counties (see Table 1 for items). As of October 2017, this empirically-derived tool was being utilized in 25 of Colorado's 64 counties.[7] CPAT scores range from 0 – 82, with four categories indicating risk to FTA and/or re-offend while released pretrial (see Table 2). The four risk categories are ranked from lowest to highest: Category 1 (scores 0-17), Category 2 (scores 18-37), Category 3 (scores 38-50), and Category 4 (scores 51-82).

Table 1. *Current CPAT risk factor and scoring*[8]

CPAT Risk Factor	Scoring	Points
Having a Home or Cell Phone	Yes	0
	No or Unknown	5
Owning or Renting One's Residence	Yes	0
	No or Unknown	4
Contributing to Residential Payments	Yes	0
	No, or Unknown	9
Past or Current Problems with Alcohol	No	0
	Yes, or Unknown	4
Past or Current Mental Health Treatment	No	0
	Yes, or Unknown	4
Age at First Arrest	This is my first arrest	0
	35 years or older, or Unknown	0
	25-34 years	10
	20-24 years	12
	19 years or younger	15
Past Jail Sentence	No, or Unknown	0
	Yes	4
Past Prison Sentence	No, or Unknown	0
	Yes	10
Having Active Warrants	No	0
	Yes, or Unknown	5
Having Pending Cases	No	0
	Yes, or Unknown	13
Currently on Supervision	No	0
	Yes, or Unknown	5
History of Revoked Bond or Supervision	No	0
	Yes, or Unknown	4

[6] Pretrial Justice Institute. (2012). *Colorado Pretrial Assessment Tool (CPAT)*.

[7] Cain, M. A. (2012, October). Summary sheet regarding Pretrial Services in Colorado. *Colorado Criminal Defense Institute.*

[8] Tables 1 and 2 were adapted from Colorado Association of Pretrial Services (2015, June). *The Colorado Pretrial Assessment Tool (CPAT): Administration, scoring, and reporting manual (Version 2.1).*

Table 2. *Current CPAT category breakdown by scores from original dataset*

Risk Category	Risk Score	Public Safety Rate	Court Appearance Rate
1	0 - 17	91%	95%
2	18 – 37	80%	85%
3	38 – 50	69%	77%
4	51 - 82	58%	51%

Methodology Summary

The purpose of this study was to assess the performance of the CPAT and identify evidence-based improvements to the assessment tool and implementation protocol. In partnership with University of Northern Colorado's Department of Criminology and Criminal Justice and the Pretrial Executive Networks (PEN), a comprehensive research plan was developed with the participation of pretrial service agencies in seven counties (Boulder, Denver, Garfield, Larimer, Mesa, Pueblo, and Weld). The following objectives and research questions were defined by the researchers to meet the study's purpose.

Objectives

1. Evaluate the relationship between the CPAT risk scores and pretrial outcomes.
2. Evaluate the predictive performance of the CPAT risk classification levels.
3. Evaluate the assignment of bond conditions in accordance with the CPAT risk classification levels.
4. Identify evidence-based strategies for effective implementation of the CPAT.

Research Questions

1. How often is the CPAT's risk level assignment consistent with certain pretrial outcomes?
2. Are the CPAT's risk levels designated, weighted, and scored for the instrument's best predictive ability?
3. Are the risk factors in the CPAT the best predictors of certain pretrial outcomes?
4. What is the perceived utility of pretrial risk assessment by those who carry out the pretrial process?
5. Is the CPAT implemented consistently and effectively?

Phases

This study intended to test the function and predictive performance of the CPAT as it is currently used in Colorado, examine stakeholder feedback regarding the tool's utility and implementation, and to create and pilot an alternative version of the tool, the CPAT-Revised (CPAT-R). To accomplish these goals, a three-phase study was designed beginning in January 2018. This section will address the methodological goals used in each section, broken down into phases.

Phase One: CPAT Validation (January 2018 – June 2018)

- Retroactive validation of the CPAT using cases from 2015 and 2016.
- Online perception survey of pretrial officers, administrators, judges, prosecutors and public defenders.

Phase Two: Analysis of Implementation (May 2018 – July 2018)

- Focus groups of pretrial officers, administrators, judges, prosecutors and public defenders about implementation.
- Onsite observation of participating pretrial agencies assessment interview and investigation processes.

Phase Three: Pilot Test (November 2018 – June 2020)

- Construct an alternative tool based on the retroactive validation and feedback from stakeholders.
- Pilot test - random assignment of the alternative tool.
- Validate CPAT-R and assess differences across sub-groups.

Outcomes

In phases one and three, similar outcome measures were used to validate the CPAT and the CPAT-R. For phase one, new arrest was identified as any new offense noted in a defendant's criminal history record, which occurred after the defendant was released on bond and prior to sentence disposition. For phase three, new arrest was identified as any new charge filing during the pretrial release time period. The state identifier (SID) was not available for the 2015-2016 retroactive data in one of the seven participating counties. This limited the court records that could be retrieved in this county and criminal history records were used instead to define the new arrest outcomes for phase one. The SID was available in all of the participating counties for the third phase of the study allowing for court records to be used to define the new arrest outcome. These records were more reliable than criminal history records at defining pretrial outcomes because they are less vulnerable to over-counting multiple arrest events for a single charge. Records from pretrial service agencies in the participating counties, the Colorado Judicial Branch, Denver Municipal Court Records system, and the Colorado Bureau of Investigation were used to construct this timeline. New arrest was also measured a variety of ways, to accurately capture the differences in the type of new arrest after bond release. New arrest was defined as:

- New arrest at all: This represents any new arrest noted in official records regardless of offense type,
- New arrest – violent: This represents any new arrest that is a victim involved, violent act (i.e., assault, domestic violence),
- New arrest – DV or order violation: This represents any new arrest that involves domestic violence as defined by Colorado statute, or the violation of a protective or restraining order,

- New arrest – serious: This represents any new arrest that involves an aggravated, felony, or violent offense, and
- New arrest – other: This represents any new arrest that is not categorized by Colorado statute to be a victim involved, violent act, or involve domestic violence.

To capture differences amongst FTAs, this pretrial outcome variable was measured four different ways. Measurement of these FTA variables was defined by the judicial response after the FTA noted in the Colorado Judicial Branch court records and Denver Municipal Court Records system. FTA is defined as:

- FTA at all: This represents any FTA noted in court records regardless of reaction by the court,
- FTA – no consequence: This represents any FTA that is followed by no response by the courts (i.e., no enhancement of bond conditions, no warrant issued)
- FTA – low consequence: This represents any FTA that is followed by a moderate sanction by the courts (i.e., bond increase, bond condition enhancement), and
- FTA – high consequence: This represents any FTA that is followed by a formal sanction by the courts (i.e., issuing a warrant).

Phase One: CPAT Validation

Part 1: Retroactive Validation Methodology

Data Sources. Records for the retroactive validation were collected from four different sources: participating pretrial service agencies, Colorado Judicial Branch records, Denver Municipal Court Records system, and the Colorado Bureau of Investigation criminal history records. Data were collected from CPAT interviews that were conducted in 2015 and 2016. The unit of analysis was the CPAT interview. Data contained information pertaining to this and the corresponding pretrial defendant. A stratified sample was collected and framed according to the number of bookings per county, per year.

Data Collection. The participating counties were initially surveyed about the records that they maintained to determine the variables that could be collected for the retroactive validation. In sum, variables that represent a CPAT interview, and corresponding pretrial defendant's demographic, bonding, pretrial supervision, and charge information were collected for the study. Court records were collected from the Colorado Judicial Branch, as well as the Denver Municipal Court Records system. These records represent a CPAT interview's corresponding case information, events, active warrants, and bond information. Finally, criminal history records were collected from the Colorado Bureau of Investigation. These records represent the pretrial defendant which corresponds to the CPAT interview's criminal history from first arrest to when records were collected.[9]

Descriptive Statistics. Overall, records for 4,600 CPAT interviews were collected from the participating pretrial services agencies. These were matched to the corresponding court and criminal history records. Matched court records resulted in a total sample of 3,386 CPAT interviews and corresponding defendants for both 2015 and 2016. A reduced sample of 3,124 was used to assess new arrest. This reduced sample is employed as a result of failed matched records. Table 3 reports the number and proportion of CPAT interview contribution per county.

Table 3. *Breakdown of participating county defendants in Phase 1 study*

County	Number of Interviews in Total Sample	Percent of Total Sample
Boulder	226	6.67%
Denver	1527	45.10%
Garfield	85	2.51%
Larimer	489	14.44%
Mesa	283	8.36%
Pueblo	445	13.14%
Weld	331	9.78%
Total	3,386	100.00%

[9] CBI CCIC records represent criminal histories from first arrest to date collected, May 21, 2018.

Research questions: The retroactive validation sought to answer three of the five full project research questions:

1) How often is the CPAT's risk level assignment consistent with certain pretrial outcomes?
2) Are the CPAT's risk levels designated, weighted and scored for the instrument's best predictive ability?
3) Are the risk factors in the CPAT the best predictors of certain pretrial outcomes?

Analysis: The key variables utilized in the validation were CPAT predictor variables (see Table 1), and variables representing pretrial outcomes including, FTA to a scheduled court event and new arrest. A variety of analytic techniques were utilized to assess the CPAT's predictive performance. To understand how well the CPAT predicts new arrest and FTA, the receiver operating characteristic (ROC) was employed. This approach estimates the probability of the pretrial outcome being consistent with the assigned CPAT risk level, such as, successful case disposition, new arrest, or FTA. To understand how the weighting and risk levels of the CPAT perform, the CPAT was compared to alternative weighted and risk level designations to estimate whether the CPAT, in its current state, performed the same or better than a possible alternative.

Part 1: Results

1) How often is the CPAT's risk level assignment consistent with certain pretrial outcomes?

Retroactive Validation

The CPAT designates four distinct risk categories ranging from 1 (lowest risk) to 4 (highest risk). Of the total sample, 11.06% were designated as Category 1, 36.91% as Category 2, 29.26% as Category 3, and 22.78% as Category 4. Table 4 reports descriptive estimates of the proportion of pretrial outcomes per type.

Table 4. *Number of pretrial outcomes the current sample*

Pretrial Outcome	***N* (%)**
New Arrest or FTA – at all[a]	1354 (43.33%)
FTA – at all	1015 (29.98%)
FTA – no consequence	202 (5.97%)
FTA – low consequence	128 (3.78%)
FTA – high consequence	685 (20.23%)
New Arrest – at all[a]	903 (28.90%)
New Arrest – violent[a]	51 (1.63%)
New Arrest – DV or order violation[a]	107 (3.43%)
New Arrest – serious[a]	293 (9.38%)
New Arrest – other[a]	841 (26.91%)

a) Estimates based on reduced sample, n = 3,124

The ROC estimates the probability of a true and false, positive and negative risk assignment. For example, a low scoring Category 1 CPAT designation that is expected to have a successful pretrial outcome (negative), but instead has a new arrest outcome (false) would be a false negative. In contrast, a low scoring Category 1 CPAT designation that does have a successful pretrial outcome (true) would be defined as a true negative. The area under the curve (AUC), which is estimated using the ROC, is used to infer the probability that the risk assessment will assign a score that is consistent with the actual pretrial outcome. An AUC score of .50 or lower means that the risk assessment score predicts pretrial outcomes no better than chance.[10] An AUC score of 1 indicates perfect predictive performance. Table 5 reports the AUC score for each pretrial outcome of interest.

Table 5. *AUC scores and confidence intervals for each pretrial outcome*

Pretrial Outcome	**AUC (BC CI)**[11]
New Arrest or FTA – at all[a]	.58* (.56 - .60)
New Arrest – at all[a]	.54* (.52 - .57)
New Arrest – violent[a]	.54 (.45 - .63)
New Arrest –DV or order violation[a]	.51 (.45 - .57)
New Arrest – serious[a]	.58* (.54 - .61)
New Arrest – other[a]	.54* (.52 - .57)
FTA – at all	.54* (.52 - .56)
FTA – no consequence	.53 (.49 - .57)
FTA – low consequence	.53 (.49 - .56)
FTA – high consequence	.54* (.51 - .56)

* = Estimate is likely beyond chance
a) Estimates based on reduced sample, *n* = 3,124
b) Confidence Intervals (bootstrap 1000)

Overall, the CPAT validates for all pretrial outcomes. This means that the CPAT was more likely to predict the actual pretrial outcome than chance alone. However, the low AUC scores across the primary pretrial outcomes does provide support for further exploration of improvements to the CPAT to increase the predictive performance.

[10] Hanley, J.A. & McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 148: 839-843.
[11] The bootstrap confidence interval (BC CI) indicates that the AUC score in the larger population is likely within this score range.

2) Are the CPAT's risk levels designated, weighted and scored for the instrument's best predictive ability?

Risk Level Designation

The four category risk level designation is assessed two ways. First, a visual assessment of the predictive curve was conducted to identify any meaningful breaks in the likely pretrial outcome. Although breaks were visualized, further examination of the possible addition of a 5th risk category was made within risk category 2. The optimal cut-point within Category 2 was identified by maximizing the sensitivity and specificity of the Category 2 risk scores. The optimal point within Category 2, which ranges from 18 – 37 points, was identified at scores 28 and 29. The predictive performance and likelihood of a new arrest and/or FTA outcome of the current CPAT Category 2 and alternative two-part Category 2 (Cat2A range: 18 – 28 & Cat2B range: 29 – 37) were compared. The likelihood of a new arrest and/or FTA outcome was also estimated across this two-part risk category. Table 6 reports these estimates.

Table 6. *Likelihood of new arrest and/or FTA and predictive performance category 2 comparisons*

CPAT Category 2	Odds Ratio (SE)	Confidence Interval	AUC
Current CPAT 18 -37	1.05* (.01)	1.03 – 1.07	.54*
CPAT 2A 18 -28	.63* (.08)	.49 – .81	.56*
CPAT 2B 29 - 37	1.58* (.21)	1.23 – 2.04	.56*

* = Estimate is likely beyond chance

a) Estimates based off sample of category 2 CPAT interviews, $n = 914$

The AUC score for the alternative two-part Category 2 is slightly higher, indicating a better predictive performance. Furthermore, when comparing the two-part Category 2, those who score at the lower half of Category 2 are less likely to have a new arrest and/or FTA outcome compared to those who score higher. This estimate indicates that there is meaningful variation in the likely pretrial outcome amongst those assigned a risk Category 2. This finding supports the need for further assessment of the risk level designation.

Weighting and Scoring

The weighting and scoring of the CPAT is assessed two ways. First, the weighting and scoring estimation used in the construction of the CPAT is replicated on the retroactive sample to infer stability of the weighting measures. To do this, the marginal effects of each CPAT item on the likelihood of a new arrest or FTA occurring are estimated. In other words, the amount of influence each individual item has on the overall risk of new arrest and/or FTA. These effects are translated into percentage points. Table 7 reports the re-estimated and current CPAT points. These weights were used to calculate an alternative weighted CPAT. The AUC was compared across this alternative CPAT and the current CPAT indicating that the current CPAT's predictive performance was comparable (CPAT current: AUC = .59, BC CI: .57 - .61), Reweighted CPAT: AUC = .59, BC CI: .56 - .61).

Table 7. *Weighted effects of each CPAT risk factor on the instrument's outcome score*

CPAT Risk Factor	Margins weight (points)	CPAT weight (points)
Have a home or cell phone	.06 (6)	Yes = 5
Owning or renting one's residence	.05 (5)	Yes = 9
Contributing to residential payments	.09 (9)	Yes = 9
Past or current problems with alcohol	.06 (6)	Yes = 4
Past or current problems with mental health	.03 (3)	Yes = 4
Age at first arrest	.08 (8)	Age <19 = 15 Age 20-24 = 12 Age 25-34 = 10
Past jail sentence	.04 (4)	Yes = 4
Past prison sentence	.01 (1)	Yes = 10
Having active warrants	.09 (9)	Yes = 5
Having other pending cases	.02 (2)	Yes = 13
Currently on supervision	.02 (2)	Yes = 4
History of revoked bond or supervision	.07 (7)	Yes = 4
Score Range	0 - 62	0 - 82

The second approach employed the Burgess method of re-weighting the CPAT items. This approach assigns a value of 1 to each "yes" response for every risk item and reduces any variation introduced in the risk score by value weighting. The AUC was also compared across this alternative CPAT and the current CPAT producing estimates that were also comparable (CPAT current: AUC = .59, BC CI: .57 - .61), Reweighted CPAT: AUC = .59, BC CI: .56 - .61). Findings support that the item weighting and scoring of the CPAT in its current state does not interfere with the tool's predictive performance when compared to alternative approaches.

3) Are the risk factors in the CPAT the best predictors of certain pretrial outcomes?

To assess whether the risk factors currently used in the CPAT are the best predictors of certain outcomes, the CPAT in its current state is also tested against an alternative tool with modified risk factors. To select meaningful risk items for the alternative tool, propensity score matching was used.[12] This approach involved matching across CPAT items to isolate the meaningful differences in the unmatched, omitted variables. Said another way, this approach tests the relationship of each omitted risk indicator to new arrest and/or FTA when all other risk indicators were the same. Iterations of this matching scheme were conducted for each CPAT item. Items identified as meaningful in this matching process include: having a home or cell phone, contributing to residential payments, and reporting a past or current alcohol problem.

[12] Features of the propensity score matching approach were: 1:1, greedy matching, non-replacement, with a caliper of .25.

A subsequent analysis was then conducted using multivariate regression to identify any other meaningful variables to include in the alternative tool: having an active warrant at the time of arrest, having a pending case at the time of arrest, and the age of first arrest. Using these variables, an alternative CPAT risk assessment was constructed that ranged from 0 – 55. To maintain consistency and allow for appropriate comparisons, the original weights of each item were used for the alternative tool. Table 8 reports the risk indicators included in the current CPAT and alternative tool.

Table 8. *Current CPAT and alternative tool risk factors*

CPAT Risk Factor	Current CPAT	Alternative Tool
Having a Home or Cell Phone	X	X
Owning or Renting One's Residence	X	
Contributing to Residential Payments	X	X
Past or Current Problems with Alcohol	X	X
Past or Current Mental Health Treatment	X	
Age at First Arrest	X	X
Past Jail Sentence	X	
Past Prison Sentence	X	
Having Active Warrants	X	X
Having Pending Cases	X	X
Currently on Supervision	X	
History of Revoked Bond or Supervision	X	

X = Risk factor included in the tool

Table 9 reports the predictive performance of the CPAT in its current state compared to this alternative tool. The alternative tool that was made up of risk items from the original CPAT is slightly better at predicting all 3 primary outcome variables: new arrest and/or FTA, FTA, and new arrest. This estimate provides support for further exploration of an alternative assessment tool.

Table 9. *Comparison of AUC scores from original CPAT to the alternative tool*

CPAT	New Arrest/FTA AUC (BC CI)[a]	FTA AUC (BC CI)	New Arrest AUC (BC CI)
Current CPAT	.59* (.53 - .61)	.59* (.51 - .59)	.55* (.52 - .61)
Alternative CPAT	.60* (.54 - .61)	.62* (.56 - .64)	.56* (.50 - .59)

* = Estimate is accurate beyond chance

a) Estimates based on reduced sample, *n* = 3,124

b) Confidence Intervals (bootstrap 1000)

Part 2: Process Evaluation

Data Sources: In an attempt to gain insight on the perceptions that multiple stakeholders in Colorado's criminal justice system have about the CPAT, a survey was constructed and disseminated to pretrial services employees, judges, prosecutors, and defense attorneys. The survey served as an information-gathering tool to further inform focus group discussion as part of phase two of this project. The survey was constructed by the researchers after numerous meetings with pretrial services supervisors, at both PEN meetings and one-on-one meetings with multiple counties. Once a draft of the survey was complete, feedback was solicited from pretrial service supervisors, with some questions modified or added.

Data Collection: The survey was created and distributed using the online survey tool Qualtrics (Appendix A). Chain-referral sampling was utilized as a technique to gain access to stakeholders throughout the state of Colorado through a top-down distribution strategy. To distribute the survey, a number of email lists were compiled: court administrators through each of the judicial districts in Colorado, District Attorney's offices in each judicial district, office heads of the public defenders in each jurisdiction, and pretrial offices throughout the PEN group, and the Colorado Association of Pretrial Services (CAPS) listserv. Surveys were distributed to these email addresses with a message requesting that it be shared with all representatives in their jurisdiction. In addition, pretrial supervisors were encouraged to share the survey link to stakeholders within their jurisdiction in an attempt to maximize the number of responses. The survey was first distributed on March 13, 2018, with reminder emails distributed April 2nd and April 12th. The survey closed on April 27th.

Descriptive Statistics: The survey yielded a sample size of 382 respondents. The breakdown of these respondents included 122 from pretrial services (31.9%), 123 defense attorneys (32.2%), 69 prosecutors (18.1%), and 68 judges (17.8%) (Table 10). There were a total of 27 out of a possible 64 Colorado counties represented in this study, with seven respondents indicating that they represented multiple counties. Denver County had the most representation with 39 respondents, followed by Mesa (n = 29), El Paso (n = 27), Boulder (n = 28), Larimer (n = 22), Weld (n = 19), Jefferson (n = 18), Pueblo (n = 17), and Adams (n = 12). The remaining counties had fewer than 10 respondents, with 117 respondents failing to identify their county.

Table 10. *Respondents' identified stakeholder role by participating counties*

County	Pretrial Services	Judges	Defense	Prosecution	Total
Boulder	11 (4.2%)	4 (1.5%)	5 (1.9%)	8 (3.0%)	28 (10.6%)
Denver	12 (4.5%)	8 (3.0%)	6 (2.3%)	13 (4.9%)	39 (14.7%)
Garfield	3 (1.1%)	3 (1.1%)	0 (0%)	2 (0.8%)	8 (3.0%)
Larimer	12 (4.5%)	5 (1.9%)	2 (0.8%)	3 (1.1%)	22 (8.3%)
Mesa	6 (2.3%)	3 (1.1%)	4 (1.5%)	16 (6.0%)	29 (10.9%)
Pueblo	8 (3.0%)	7 (2.6%)	1 (0.4%)	1 (0.4%)	17 (6.4%)
Weld	11 (4.2%)	6 (2.3%)	2 (0.8%)	0 (0%)	19 (7.2%)
Other	59 (15.4%)	32 (8.4%)	103 (27%)	26 (6.8%)	220 (57.8%)
Total	122 (31.9%)	68 (17.8%)	123 (32.2%)	69 (18.1%)	382 (100%)

Note. The "Other" category includes those from any county not participating in this study as well as any participants (n = 117) who did not identify their county.

Part 2: Results

Familiarity with CPAT Tool

Among the first items of the survey, respondents were asked to score their familiarity with the CPAT (range 1 – 10). A score of one indicated the respondent was not familiar with the tool at all. A score of five indicated an understanding of the instrument conceptually, with a ten being extremely familiar. In comparing the four criminal justice roles among their familiarity scores, a One-Way ANOVA was estimated. Table 11 displays the average familiarity scores among the four groups. The One-Way ANOVA indicates statistically significant differences among two of the groups, $F(3,368) = 4.27$, $p = .006$, as pretrial officers rated higher levels of familiarity with the tool when compared to judges ($p = .03$) and prosecutors ($p = .01$).

Table 11. *Familiarity with the CPAT: Mean scores by criminal justice role*

Role	N	Familiarity Mean Score (*SD*)
Pretrial Services	119	8.71 (1.64)
Defense Attorneys	118	8.21 (1.70)
Judges	67	7.97 (2.00)
Prosecution	68	7.88 (1.77)
Total	372	8.26 (1.78)

Note: Range = 1-10; Ten respondents failed to answer this question.

Each respondent was asked questions relating to how they felt about the CPAT in its current state, with follow-up questions further specifying what they liked about the CPAT and what they felt could be improved. Pretrial officers generally had favorable comments regarding the CPAT in its current state, with a few respondents reporting distaste for the tool as it is currently being used. One respondent exemplified this point in stating:

> *I find it concerning that [the CPAT] has been used for years, and there are still many questions and debates about how to score certain questions. It seems that a validated, simplified tool would be best to use as a statewide tool to increase implementation consistency…* (Participant 110. Pretrial).

Perceived Issues with Current CPAT Tool.

Thematic analysis was employed through all survey respondents' qualitative, written responses. Results showed seven key themes emerged relating to the issues with the CPAT in its current form.

Prior Failure to Appear: All four groups expressed concerns with the lack of prior FTAs being considered in the score for the CPAT. As the tool is a risk assessment that predicts future FTAs, the lack of previous FTAs within the 12 items compromised its face validity.

Substance Abuse item: Another common concern among the four groups involved the lack of substance abuse questions. While there is a self-report item addressing problematic alcohol use, narcotics are not included as a risk factor in the current CPAT.

Reliability of Self-Report Measures: The self-report risk factors on the CPAT have raised concerns from all four groups of stakeholders. While some jurisdictions work to verify information provided as part of the CPAT interview, others do not have the resources. The perceived overreliance on self-reporting among some respondents resulted in the belief that defendants are rewarded for providing false information. As one judicial respondent noted:

> *I fear that defendants and defense counsel can 'game the system' by not being truthful in the assessment. Since many of the questions asked to determine the score depend on the defendant's truthfulness, a defendant need only fib a little to skew his/her score. I've seen a few examples of this. In my view, the more we can rely upon objective reporting of the information underlying the defendant's score, the better* (Participant 123, Judge).

Category 2 Overload: Among pretrial officers, a commonly identified issue relates to the second classification category. Defendant's CPAT scores classify them into four discrete categories, based on their risk to FTA and reoffend. Many pretrial respondents identified that the second category is too broad, causing most individuals to score in the second category. This sentiment was also found in the judge's responses, with multiple judges emphasizing that most defendants fall into either Category 2 or 3.

Definitions: Some items were unclear to the respondents, such as what constitutes as "mental health problems" and "contributing to residential payments." Many defense attorneys felt that the tool unfairly penalizes indigent defendants. As a number of items penalize defendants for not having a phone, renting instead of owning, and not contributing to residential payments, those with lower or no income are scored higher than those who have more economic advantages.

Crime Type, Severity, and Risk: Respondents from the judicial, defense, and prosecution groups recommended that the severity of the crime be considered in a risk assessment tool. In addition, a number of pretrial respondents identified that the CPAT should differentiate the risk to reoffend from the risk of failing to appear in court. In its current form, the CPAT combines these two risk indicators into an overall risk score.

Buy-in: When asked how support of the tool could be improved among judges, prosecutors, and defense attorneys, respondents overwhelmingly stated that more education would help each group adhere to the protocol of the CPAT. A clear explanation of the purpose and goals of risk assessment tools could help each group understand what the CPAT is meant to and not meant to do. One pretrial respondent succinctly summarized the narratives surrounding the buy-in of the CPAT in the following quote:

> *In short, it is a good tool. Much of the criticism around the tool comes from a fundamental misunderstanding of the purpose, use, value, and limitations of assessment by our system* (Participant 73. Pretrial).

CPAT Implementation

There were a total of 95 respondents (28.9%) who conduct CPAT interviews as part of their job. This group included 73 pretrial services staff, 19 defense attorneys, and three judges throughout the state. Eighteen of the 27 counties that participated in the survey had individuals who reported conducting CPAT interviews as part of their job. Within those counties, 50% of the respondents reported that their CPAT reports are reviewed by a second individual before being submitted to the court. When broken down by county, eight of the eighteen counties reported that they do not have another individual review their reports.

CPAT Tasks and Importance

Those who identified administering the CPAT as part of their job duties were asked to rank the importance of the five CPAT tasks from most important (rank = 1) to least important (rank = 5). Table 12 displays the mean ranks of these tasks. The interview was ranked as the most important (M = 1.91, SD = 1.04), followed by criminal history checks (M = 2.04, SD = 1.10), confirming information (M = 3.01, SD = 1.01), the CPAT assessment score (M = 3.33, SD = 1.15), and victim interviews (M = 4.71, SD = 0.56), respectively. The importance of these ranks were defined as most valuable in achieving a reliable CPAT score or reliable information about a defendant.

Table 12. *Ranked important of CPAT tasks by those who administer the CPAT (n = 76)*

Rank	Task	Mean Rank (*SD*)
1	Interview	1.91 (1.04)
2	Criminal History Checks	2.04 (1.10)
3	Confirming Information	3.01 (1.01)
4	CPAT Assessment	3.33 (1.15)
5	Victim Interviews	4.71 (0.56)

Note: Of the 95 respondents who conduct CPAT interviews, only 76 answered this question

The majority of respondents from all four stakeholder groups reported the interview as a vital part of the CPAT process. Two-thirds of respondents (66.9%) reported that they would not support a risk assessment tool if it did not involve an interview with the pretrial defendant. Respondents from all four groups claimed that the interview component was important to gather information that could not be obtained from official records. The perceived drawback is that

many defendants have the opportunity to provide inaccurate information during interviews. As one pretrial officer responded:

> *The interview is important… as civilian personnel we can get more information than Law Enforcement Officers are able to. The pitfall of the interview is that the defendants can say whatever they want, which can lead some people to not tell a single true fact about themselves (rendering the interview somewhat useless). The verification portion is thus important (though the same thing applies to verifiers: they can say whatever they want, even if it is not the truth)*. (Participant 93. Pretrial)

Those who believed that the interview was not necessary for a pretrial risk assessment tool noted the bias that could be derived from an interview. As one defense attorney noted:

> *In my experience, the interview tends to be highly subjective and falls prey to the biases of the interviewer, who is often unsympathetic to indigent defendants. Where I practice, the interviewer is also who the defendants must check in with for pre-trial tracking services. That interviewer often carries biases from the interviews into his supervisory role and tends to "play favorites" with clients, leading to reduced credibility with the court and unequal treatment of defendants.* (Participant 222. Defense)

To combat the bias of the interview, all respondents overwhelmingly agreed that confirming information with outside sources is important. The most common response to why confirming information is important in the CPAT process lies in the perceived integrity of the defendant. As noted above, many respondents reported that defendants can benefit from giving inaccurate information in order to reduce their CPAT score.

Phase 2: Analysis of Implementation

Methodology

Focus Groups

The focus group component of the phase two analysis addresses the broad research question: *What is the perceived utility of pretrial risk assessment by those who carry out the pretrial process?* A total of 14 focus groups were conducted between May, 2018 and June, 2018. Focus groups were held with two categories of individuals involved in the pretrial process: a) pretrial officers who conduct risk assessment interviews and investigation, as well as pretrial supervisors, and b) pretrial stakeholders who use risk assessment tools to inform release decisions and bond arguments (e.g., judges, prosecutors, defense attorneys, and other criminal justice administrators). In five of the six counties, one focus group was conducted with pretrial officers and supervisors and another with pretrial stakeholders. In the sixth county a total of four focus groups were conducted. The four focus groups in one county were the result of the large number of interested participants in certain roles in the county. To ensure that each role amongst the large number of participants was represented in the feedback, focus groups were conducted for each role. Participants for each of these four focus groups were defined by role in the pretrial process: pretrial officers and supervisors, judges, prosecutors, as well as defense attorneys.

Participants were recruited through two methods. First, the phase one survey of criminal justice stakeholders ended with an inquiry into their interest in taking part in a focus group in the future. This survey was distributed via various listsevs of state agencies that have direct roles in the pretrial process (i.e., pretrial officers, prosecutors, judges, public defenders) and through chain-referral sampling of those who had completed the survey. If interested, these participants provided an email address. The researchers later invited these individuals to take part in a focus group discussion. Due to low availability or non-responses to these emails, a second strategy was employed. This second strategy involved pretrial administrators in each of the participating counties recruiting participants through county-wide email listsevs. During the one-hour focus groups, one of the authors served as lead-facilitator, one as co-facilitator, and at least one other serving as a note taker. The focus groups were audio recorded and followed a semi-structured interview guide that will be further discussed below.

In total, 109 participants took part in the 14 focus groups. Six of the focus groups were with pretrial officers and supervisors, involving 41 individuals. Eight of the focus groups were with pretrial stakeholders (i.e., judges, prosecutors, defense attorneys, jail staff) involving a total of 68 individuals. To ensure confidentiality, demographic information was not collected about the focus group participants. Each focus group lasted about one hour and was audio-recorded. These recordings were later transcribed and de-identified to further ensure confidentiality of the participants.

Observations

From May 2018 to June 2018, onsite observations were conducted at all seven participating counties. These observations were similar among jurisdictions, typically involving the researchers shadowing pretrial officers for one morning. These observations allowed for the researchers to understand how the CPAT was being administered throughout the seven participating counties and track any differences in the tool's implementation, scoring, and use in advisement hearings.

While no systematic analysis was conducted for the observation phase of this study, the researchers observed differences across pretrial agencies in how the CPAT is administrated and scored. Each of the seven counties had varying space within their jails to conduct CPAT interviews. This ranged from full office space in the booking section of the jail to conducting video interviews with defendants from an office location outside of the jail. Three counties did not have any designated jail space for pretrial services within the jail's central booking locations. These counties had to rely on private meeting rooms for interviews. Overall, the recommendations from these onsite observations call for uniform implementation, or standardization, of a pretrial risk assessment tool for all participating counties. This could be accomplished with standardized training for all pretrial staff.

Results

A semi-structured interview protocol was created for each type of focus group (pretrial officers and supervisors, as well as judge, prosecutors, and defense attorneys) (Appendix B). This allowed for the facilitators to pose probing follow-up questions to guide the discussion and solicit in-depth and different perspectives. The interview protocol differed slightly between the two focus groups due to the difference in familiarity with the administration of the tool between pretrial officers, supervisors and pretrial stakeholders. The initial questions were broad and included how participants felt about the tool in its current state and the utility of the tool for them. Other questions pertained to perceptions of buy-in for the tool, training, and how or if their views of the tool would be impacted if certain modifications were made. The interview protocol was followed for all of the focus groups. There were instances that focus group conversations would stray from the protocol; however, such conversations still pertained to the pretrial risk assessment tool in question, and thus were welcomed by the facilitators.

Each focus group was recorded, transcribed, and transcripts were then inputted into a qualitative data analysis software, NVivo. Thematic analysis was conducted upon coding each focus group transcription applying Braun & Clarke's (2006)[13] recommendations for conducting thematic analysis. Emergent themes were identified by reviewing each transcription by the researchers analyzing the transcripts and coding similar comments into nodes. To address inter-rater reliability between the coders, reliability checks of coding and themes were conducted through peer debriefing. Any discrepancies in themes were discussed and clarified among the researchers. Once identified, the researchers reviewed each of the NVivo coded responses and transcripts to confirm these themes and identify quotations that exemplify these themes.

[13] Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology. 3*(2), 77-101.

Five themes emerged from the thematic content analysis. These themes address the study research question: What is the perceived utility of pretrial risk assessment by those who carry out the pretrial process? The themes include: 1) the role of the risk assessment tool, 2) risk override and discretion, 3) informing pretrial supervision and outcome, 4) consideration of other factors, independent of the risk assessment, and 5) training and education for tool.

Theme 1: The Role of the CPAT

The role of the CPAT in the pretrial process was found to be dependent on the perception about how these tools are used. Specifically, how this tool is used when informing the bond release decision. This is especially relevant to the implementation and resulting accuracy of an assessment tool. Prior research has shown that negative perceptions of risk assessment tools can result in pushback about adherence to implementation protocol.[14] Diverting from implementation protocols can ultimately compromise the accuracy of an empirically constructed and validated assessment tool.[15]

Feedback from both categories of focus groups, pretrial officers and supervisors, as well as pretrial stakeholders indicated that the CPAT played a role in the arraignment process. Participants were in agreement that the role of these tools was to accompany professional discretion. One pretrial officer noted: "*I want a tool to be used as a foundation…. use your professional judgement…it's a starting point, it's an anchor*" (Pretrial participant: Pretrial Officer).

Importantly, many of the respondents did not prioritize the CPAT in their decision-making process. The pretrial risk assessment score and category are also perceived by stakeholders as being no more or less important relative to other factors that are considered. Rather, the risk score and category are perceived as one piece of information amongst many that are considered. A pretrial stakeholder describes its role:

> "*I think the best way to describe it is, from our perspective, it's a piece of information. It's not something that we are heavily relying upon, in making our arguments to the judges because we are still going to have to go back and do the work*" (Stakeholder participant: Prosecutor).

[14] Gottfredson, D., Gottfredson, S., & Conly, C. (1989). Stakes and risk: Incapacitative intent in sentencing decisions. *Behavioral Sciences and the Law, 7,* 91-106.

[15] Latessa, E., & Lovins, B. (2010). The role of offender risk assessment: A policy maker guide. *Victims and Offenders, 5*(3), 203-219.

Theme 2: Risk Override and Discretion

Another emergent theme from the thematic content analysis is the difference in how the CPAT risk score and risk decided on professional discretion alone were perceived. Overriding a risk score as an exercise of professional discretion was especially relevant to judges who are accountable for public safety. Judicial override occurs when a judge decides not to abide by the recommendation of the pretrial officer's bond report, which is often based on a combination of the risk assessment category and state laws (i.e., Victim's Rights Act). These decisions require a balance of their exercise of professional discretion and the challenge of predicting a pretrial outcome.

This balance can be a challenge when professional discretion and a risk score do not agree. Based on feedback across all the stakeholder focus groups, the risk assessment score was thought to occasionally differ from risk decided on professional discretion alone. This difference could occur due to many factors including but not limited to: how the individual interpreting the CPAT score would use it, training, what other information was provided about the defendant, and the decision-maker's perception of the defendant. Common protocol across jurisdictions was for low risk pretrial defendants to receive a personal recognizance (PR) bond unless the judge decided to override the risk score. In numerous focus groups, the example of a sex offender was identified as a circumstance where the judge would likely override a low risk score and not order pretrial release on bond. Prosecutors and defense attorneys who utilized the pretrial risk assessment score to inform bond arguments also used the tool differently according to the specific case. As one judge describes:

> "*So the [defense attorneys] are always arguing for bond, and the [prosecutors] are all arguing for no bond…And if the CPAT is in their favor they argue it, and if it's not in their favor they disregard it. And not every single argument, but…often, and it's completely aggravating, because…well, I'm not getting any help from the litigants*" (Stakeholder Participant. Magistrate Judge).

Since judges used the CPAT score along with a variety of other factors to inform release decisions, the resulting release decisions can vary greatly. As one public defender noted: "*The same person could see three different judges and get three different results, based on the CPAT*" (Stakeholder Participant: Public Defender). Participants perceived this variation in the resulting decision to be negative but still supported accompanying pretrial risk assessment with professional discretion for release decisions.

Theme 3: Supervision/Outcome of the Risk Score

Supervision/outcome of the risk score refers to the variety of decisions that assessment tools were used to inform. Pretrial risk assessment tools are constructed to assess the risk of new arrest or FTA. Correctional needs are not an outcome included in the predictive performance of pretrial risk assessment tools. This means that little is known about the accuracy of pretrial risk assessment tools when used for bond condition assignment.

Across both types of focus groups, participants expressed that the pretrial risk score was used to both inform decisions about pretrial release, as well as the conditions of supervision that may be ordered if a defendant is released. All the participating counties in this study had a supervision matrix that incorporated the pretrial defendant's risk category. This matrix was used to help determine the level and type of bond supervision a defendant would be assigned during the pretrial process.

Similar to bond release decisions, the role of pretrial risk assessment to inform recommendations and orders of pretrial supervision conditions was accompanied by professional discretion. Although the pretrial risk assessment tool was not constructed to directly inform the pretrial supervision decision, the tool was favored by pretrial officers because it was thought to provide tangible information about the pretrial defendant that is useful for supervision purposes. One pretrial officer that supervised pretrial defendant's bond compliance advised:

> "*We use our CPAT scores for supervision, also. I don't think it was created to do that, but, we just use that on the supervision side to say, we think this person is a Category One, so then we set their frequency for testing, or whatever it is, their check-ins, based on that category. And then case managers have the room to flex that…depending on how they are doing…again, I don't think the [risk assessment] was intended to ever focus on supervision, we just do that because it gives a good baseline on how to supervise somebody*" (Pretrial Stakeholder: Pretrial Supervisor).

A common theme across participants was frustration about the uncertainty of the pretrial outcome of release and bond condition decisions. Pretrial risk assessment tools were reported to provide valuable information about a pretrial defendant and insight into uncertain pretrial outcomes. One pretrial stakeholder noted:

> "*Once we get our score, it's kind of just like, okay, now it is on me to decide what this guy's going to have to do, so…once we have the score, what do we do with it? What do we recommend, and what conditions are going to make this person more successful? I mean, there is really no research, or anything, that we have saying that…random drug testing isn't going to help this person, but we are just ordering it for everyone, because we don't know*" (Pretrial Stakeholder: Pretrial Supervisor).

Theme 4: Consideration of Other Factors

A theme emerged pertaining specifically to the other factors considered but not included in the CPAT. Participants in all roles emphasized the importance of considering other factors in conjunction with the risk score during the bond release decision. In the participating counties, judges were given information from pretrial services about a pretrial defendant's drug history, employment, and prior number of FTAs. This information was commonly reported on a bond report along with recommendations about release and supervision. These items were independent of the risk assessment tool and therefore not included in the resulting risk category. The type of offense (e.g., sex offense, domestic violence) was a primary consideration for release and supervision decision that is not included in the risk assessment score. If a defendant is charged with a high profile, sexual offense and is assessed to be low risk, a judge will often be hesitant to make the decision to release on bond.

Pretrial stakeholders across focus groups described how they incorporate additional information such as type of charge into their decision-making process. For example, one judge notes:

> "*From a judicial perspective, in order to maintain a level of consistency, I have to give the same weight to the CPAT that I would give on… a SAOC, or sex assault on a child, that's a category 1… but I also have to take into consideration the nature of the offense…and the nature of the history, the type of offenses they were charged with, not just plead to, and how recent in time those were. That's something I do independent of the score, because the score doesn't, really, take that into consideration*" (Stakeholder Participant: Judge, 1st and 2nd Advisement Court).

The type of charge, as well as perceived severity of that charge within the surrounding community were perceived to play a role in the pretrial release process. Charge severity and community ties were additional factors, outside of the CPAT risk items, that played a role in the stakeholder's decision-making process. One prosecutor succinctly summarized this point stating: "*Yeah I think our judges or our players recognize that the CPAT category is just a guideline…Our community values are going to be taken into consideration... high stakes and low stakes crime*" (Stakeholder Participant: Prosecutor).

The impact of the defendant's release on the local community was another aspect that was reportedly considered. These factors included the impact of pretrial release and supervision decisions on the overall jail capacity, as well as the decrease in bail industry involved cash bonds. As one stakeholder noted:

> "*One observation I have too… is that my contact…my lawyer contact with bail bonds people plummeted since we had the CPAT… they used to be around the courts and the courthouse and we haven't seen that nearly as much…*" (Stakeholder Participant: District Attorney's Office).

<u>Theme 5: Training and Education for Tool</u>

A common theme that emerged from all of the focus group discussions related to training and education. Training of risk assessment has been linked to enhanced buy-in from those who use it which carries implications for adherence to implementation protocol.[16] The respondents overwhelmingly reported that training about pretrial risk assessment was perceived to be very important. Formal training for the CPAT was provided to pretrial officers and supervisors throughout the state on a semi-annual rotation but not to the other pretrial stakeholders.

[16] Latessa, E., & Lovins, B. (2010). The role of offender risk assessment: A policy maker guide. *Victims and Offenders, 5*(3), 203-219.

One advantage of the formal training was to ensure consistency in the tool's implementation. It was reported that counties implemented and even scored certain items of the tool differently. For example, one item on the risk tool asked defendants to self-report a problem with substance abuse. As part of the instructional guide, the self-reported response to this question should be relied on. However, some counties allowed for their officers to override this item. For example, one pretrial officer noted:

> "*Some people agree, some people disagree, but if you have five DUI convictions, and you say you don't have a problem with alcohol…I think you might be a little…wrong on that. So I mean, we take it….case by case…I mean, we don't do it…sparingly we just…if it shows they have an alcohol problem, even though they say no…we'll override that and say yes*" (Pretrial Participant: Pretrial Officer).

Due to the limited amount of trainings offered, most of it was conducted within particular counties for pretrial officers. Stakeholders identified that they had not received the formal state-wide training for the tool. Many of the stakeholders supported the idea of going through training about the use of the risk assessment tool if it were more widely available. As one judge notes:

> "*I think one thing that is very important, as I think is very obvious just from the questions I've been being asked, or I've asked, is that [training] needs to be, I think, on-going. Either annually, or something, or if there is any slight modifications that we continue…because I think if you don't actually understand the instrument, or the theories, and the evidence based theories behind it, it's hard to have a lot of confidence*" (Stakeholder Participant: County Court Judge).

The content of the formal training given to pretrial officers state-wide pertains to the assessment interview and investigation protocol, how to interpret each defined risk item, and the scoring protocol for the tool. Pretrial officers and supervisors, as well as pretrial stakeholders, expressed a perceived need for more education about how the tool is implemented, as well as pretrial risk assessment in general. One stakeholder advised: "*There needs to be a lot more education, with the whole system, about what risk assessments do, what their purpose is, what their limitations are, what they are effective at, and what they're not effective at*" (Stakeholder Participant: Pretrial Administrator). It was expressed across multiple focus groups that expanding the frequency, accessibility, and content of the formal trainings and education could improve the tool's utility for informing pretrial decisions.

Often, the discussion surrounding buy-in for the use of the tool was incorporated into the discussion about training and education. Importantly, participants noted that training alone would not likely increase stakeholder buy-in of the tool. Education regarding the construction and statistics used to create and score the tool would be helpful in understanding the utility of the risk assessment. The participants perceived this level of transparency about the pretrial risk assessment tool as critical for on-going support of the use of the tool. As one stakeholder noted, "*I really think just understanding why…why these questions were the ones picked, why this works, because I feel like there's just not a lot of faith in the tool as it is. Especially with the other stakeholders*" (Pretrial Participant: Pretrial Officer). Furthermore, multiple pretrial officers and stakeholders identified that the training and education should clearly outline the role of a risk assessment tool in the pretrial process.

Phase 3: Pilot Test

Methodology

The first goal of the phase three analysis was to assess the predictive performance of a modified version of the CPAT, the CPAT-R. To do this, the CPAT-R was pilot tested in the seven participating counties. Predictive performance was assessed according to how precise the tool discriminates pretrial risk and accurately predicts the actual pretrial outcome.

To meet the goals of the phase three analysis we revisit the research questions posed for the phase one analysis. To inform the evidence-based recommendations for this study we seek to answer in this phase:

1) How often are the CPAT-R, CPAT and CPATR-SV's risk level assignment consistent with the actual pretrial outcome?
2) Are the CPAT-R's risk factor's weighted for the tool's best predictive performance?
3) Are the risk factors in the CPAT-R the best predictors of certain pretrial outcomes?

The best performing tool was then examined for accuracy equity and predictive parity across defendant race/ethnicity, gender, and residential status. Accuracy equity is defined as similar predictive performance across groups of a risk assessment score.[17] This was assessed by comparing the tool's ROC and corresponding AUC for meaningful differences. Predictive parity was defined as similar classification of high and low risk by the assessment tool across different groups. This was assessed by comparing the error of the assessment tool across sub-groups. Finally, the relationship between sub-group membership and predictive performance was assessed to identify if predictive performance of the tool was mediated by sub-group membership. The estimates from these analyses were used to identify the best performing and error balanced assessment tool.

The second goal of the phase three analysis was to assess the reliability and implementation features of the best performing and error balanced assessment tool. Reliability was assessed by the agreement between self-reported responses to assessment interview questions and assessment scores cross-referenced with official criminal history and court records. The implementation features that were assessed pertain to the risk factor definitions, and confirming reference information during the assessment investigation.

The findings from the reliability and implementation assessment were used to identify the recommended tool. This version of the assessment tool was then re-analyzed on a test dataset. Finally, descriptive statistics were analyzed using the full pilot sample about the recommended assessment tool.

[17] Berk, R., Heidari, H., Jabbari, S., Kearns, M. & Roth, A. (2018). Fairness in criminal justice assessments: The state of the art. *Sociological Methods & Research.* 1-42.

Construction of Colorado Pretrial Assessment Tool-Revised (CPAT-R)

Based on feedback from the first two phases of this project, the researchers determined that the construction of a revised risk assessment tool was warranted to improve upon the CPAT, both in its predictive ability, but also to account for feedback from criminal justice stakeholders to allow for a higher level of buy-in throughout the criminal justice system. To draft the modified tool variables relevant to pretrial outcomes were identified using the phase one retroactive data. These variables were informed by the feedback received in phases one & two, the existing literature on predictors of pretrial outcomes, and those factors currently included on the CPAT. The variables used for drafting the pilot CPAT-R pertained to: criminal history, community stability, offense type/severity and prior court appearance history (Appendix C).

Machine learning, LASSO, and random forest regression were used to identify and order the risk factors for the pilot CPAT-R. Research indicates that these approaches are useful for identifying parsimonious models amongst collinear variables.[18] Assessment tool produced from this approach will contain a relatively small number of risk factors selected from a large number of risk factors that are all similar in how they predict pretrial success or failure. The risk factors selected for the piloted CPAT-R were those that were still predictive of new arrest and/or FTA after accounting for these similarities amongst risk factors. Once these risk factors were identified, they were rank-ordered according to predictive performance and weighted on a 3-point scale. Risk factors contributed more to the pilot CPAT-R's predictive performance were weighted as 2 or 3 points and those contributing less a 1 or 2.

Table 13 details the risk factors and weighting of the piloted CPAT-R. Internal validation of the pilot CPAT-R's was conducted on a randomly selected test dataset from the phase one retroactive data. These estimates indicate that the pilot CPAT-R has better overall predictive performance compared to the CPAT in its current state (Appendix D).

[18] Berk. R. 2012. *Criminal justice forecasts of risk: A machine learning approach*. Springer Science & Business Media.

Table 13. *Pilot 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).
Time at current residence	0/1	Self-reported time living at current residence. (0 = 1 year or longer, 1 = less than 1 year or no residence).
Problems with alcohol or drugs	0/1	Self-reported current or past 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 yoa or older, 1 = 20 yoa 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).
Prior violent arrest	0/1	Prior violent arrest (0 = none, 1 = at least 1).

a) Range 0 – 22

A second version of the pilot CPAT-R assessed did not require an interview. The same process was implemented for constructing the pilot CPAT-R but using only the most predictive risk factors that could be confirmed with official criminal history and court records. Table 14 details the risk factors for the non-interview modification of the tool, the CPATR-SV.

Table 14. *Pilot CPATR-SV*

Risk Factor	Score	Definition
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).
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 - 16

Pilot of the CPAT-R

The pilot study spanned 3-months (November 13, 2018 – February 14, 2019) and was employed in the seven participating counties (Boulder, Denver, Garfield, Larimer, Mesa, Pueblo, and Weld). The primary objective of the pilot was to test the predictive accuracy of the Colorado Pretrial Assessment Tool (CPAT), as compared to a modified version, the CPAT-R. The CPAT-R is designed to assign risk of both FTA and re-arrest during the pretrial period. The pilot comparison groups included defendants who received only the CPAT assessment, and defendants who received both the CPAT and CPAT-R assessments (Appendix E).

The risk assessment used for each case was randomly assigned. Each participating site was provided a treatment assignment list. This list informed pretrial officers whether they will administer the CPAT-R in addition to the original CPAT. If CPAT-R was to be administered, pretrial officers would first administer the CPAT-R followed by the CPAT. If the defendant was unavailable for an interview, regardless of treatment assignment, pretrial officers were instructed to fill out the interview-less items of the CPAT-R (i.e., items 4-11 in on the CPAT-R pilot form). Five of the participating counties used the recommended pilot form. Two counties, Denver and Mesa, adjusted their data collection methods and built the CPAT-R into their general pretrial interview protocol. Denver provided CPAT-R data were collected for all pretrial interviews during the pilot period and cases were randomly selected for the study. After the pilot concluded, CPAT-R data were hand entered by the research team. Pretrial officers were also provided a pilot survey and implementation guide (Appendix F). The pilot implementation guide included instructions for administering the CPAT-R assessment questions, definitions, and recommended question wording.

Data Sources. Records for the CPAT-R pilot and validation were collected from five different sources: the CPAT-R pilot data, additional data from participating pretrial service agencies, Colorado court records, Denver Municipal Court Records system, and the Colorado Bureau of Investigation criminal history records. CPAT-R data were collected from the CPAT-R pilot period from November 13, 2018 to February 13, 2019. Other pretrial data were collected from pretrial agencies one year after the pilot was completed. These data contained information pertaining to pretrial outcomes and allowed for a one-year follow up period for the closing of cases.

Data Collection. Participating pretrial agencies provided data from variables including the CPAT interview and corresponding pretrial defendant's demographic, bonding, pretrial supervision, and charge information were collected for the study. As with phase one, court records were collected from the Colorado Judicial Branch, as well as the Denver Municipal Court Records system. These records match defendants to their corresponding case information, events, active warrants, and bond information. In addition, criminal history records were collected from the Colorado Bureau of Investigation.

The data were collected from multiple sources and matched on individual identifiers. Pretrial records were provided by the pretrial agencies of the participating counties with information about the pretrial defendant, release decision, bond and supervision. Court records were provided by the Colorado Judicial Branch and Denver Municipal Court. Criminal history records were provided by the Colorado Bureau of Investigation. Records were matched on individual identifiers including the SID, FBI number, pretrial number and date of birth.

The outcome variables used in phase one and three are conceptualized the same but operationalized differently. The variable definitions applied in phase one for the pretrial outcomes are the same (see "Outcomes," pg. 15). For example, a new arrest and/or FTA outcome is defined as at least one new arrest or FTA occurring during the pretrial release time-period according to official arrest records. For the phase three analysis, this variable is measured using new court filings and not official arrest records. It was determined that court records of new charges filed were less vulnerable validity threats as a result of entry error.

A new arrest is defined as charges filed for a new offense during the pretrial release time period. This time period spans the initial pretrial release to either case disposition. FTA is defined as a documented failure to appear to a scheduled court date during the pretrial release time period. This variable is also measured using court records.

Descriptives: Table 15 reports the descriptive statistics for those released pretrial for the full sample. The initial pretrial release was determined using pretrial agency and court records. Only those that were released pretrial were include in the phase three analysis because only those pretrial defendants had an opportunity for a pretrial release outcome.

Table 15. *Full sample – Released pretrial defendants*

Average Risk Assessment Score	Average Risk Score *(SD)*
Pilot CPAT-R (range 0 – 22)	9.72 (5.39)
CPAT (range 0 – 78)	38.34 (16.35)
CPATR – SV (range 0 – 16)	6.88 (4.60)

n = 3,757

Table 16. *Full sample – Released pretrial defendant outcomes*

Outcome	N (%)
Primary	
New Arrest and/or FTA	1,626 (35.59%)
New Arrest	740 (19.70%)
FTA	851 (22.65%)
Secondary	
New Arrest – Violent	101 (2.69%)
New Arrest – DV or order violation	158 (4.21%)
New Arrest – Serious	478 (12.72%)
New Arrest – Other	180 (4.79%)
FTA – No consequence	642 (17.09%)
FTA – Low consequence	114 (3.03%)
FTA – High consequence	274 (7.29%)

n = 3,757

For the full sample, the average risk score falls in the lower half of the total score range. About 36% of the full sample had a new arrest and/or FTA during the pretrial release time period (Table 16). Disaggregating the combined outcome, 19% of pretrial defendants were arrested and 23% failed to appear to court. The secondary outcomes are defined by the charge for a new arrest charge or the degree of consequence for FTA (see "Outcomes," p. 15). The most common of these more specific pretrial release outcomes are serious new arrests occurring for 12% and no consequence FTA's occurring or 17% of the full sample.

Pilot Feedback Survey

After the pilot study was completed an online survey was disseminated to the pretrial staff who took part in the CPAT-R pilot via the agency administrators (Appendix G). Similar to the phase one online survey, the pilot feedback survey was created using the online survey platform, Qualtrics. The survey included questions about the officer's use of the modified tool and their general perceptions about the tool. The survey opened February 21, 2019 and closed March 21, 2019. In addition to the initial recruitment email, a follow-up recruitment email was sent out on March 7, 2019.

There were a total of 47 individuals who completed the feedback survey on the CPAT-R pilot. 52 respondents had started the survey, but five were incomplete. Of these 47, 40 (85%) indicated that they conducted CPAT-R interviews during the pilot phase while 7 (15%) did not. Of the respondents, 48.9% worked as either a pretrial officer, pretrial specialist, or court services specialist, 12.8% were supervisors, 6.4% bond commissioners, with the rest (31.9%) either failing to indicate their position or falling into an "other" category (e.g., jail staff, probation officer, etc.). Respondents were from all seven participating counties, but more heavily represented by Larimer, Weld, and Garfield counties.

The overall feedback from the survey about the CPAT-R was positive. While some adjustments to the wording and definitions of some risk items were considered in the final implementation guide, respondents overall believed the tool was similar to the current CPAT in its use, and stronger than the CPAT in its face validity.

When asked, on average, what percentage of the pretrial risk assessment interviews that the respondent collected were CPAT-R interviews, responses ranged from 1-100%, with an average of 54.05% (*SD* = 31.49). The CPAT-R interview was report to take an average of 7.86 minutes (*SD* = 5.72) to complete. When asked about the recommended wording of questions in the CPAT-R implementation manual, 33 respondents (84%) reflected that they had used the recommended wording guide. Adjustments to the recommended wording were used to speed up the interview process or to add more explanation to questions defendants did not understand.

Regarding the confirmation of assessment responses with an external reference, only 63% of respondents attempted to contact a reference. Of those who attempted to contact a reference, contact was established an average of 27% (*SD* = 29.5) of the time. When asked their thoughts on confirming information, most respondents reported that the process is rarely able to be successfully completed. Some respondents noted that due to the volume of defendants, they are not able to confirm information at all. Others remarked that defendants often do not know the number of an individual to confirm their information, often do not provide accurate contact information, or no one responds when contacted. The information relating to criminal history data was reported to be more accurate than what is self-reported by defendants due to their inability to remember the information.

Furthermore, pilot feedback survey respondents were asked if there were risk items on the CPAT-R that were difficult to solicit from the defendant during the interview process. Most of the feedback addressed the criminal history-related information that defendants were asked to recall. Questions relating to number of arrests, number of FTAs, or prior violent arrests were difficult to answer. As one pretrial officer stated, "*not sure interviewees knew what constituted as 'violent*.'" Another respondent noted that, "*defendants have a hard time remembering and/or lie about their criminal history*."

Un-willful FTA: Stakeholders across counties anecdotally advised that FTA's recorded in court records were often "un-willful." Un-willful FTA's were described as those that occur at no fault of the pretrial defendant. A common example provided was when a defendant missed a scheduled court appearance because they were in custody in another jurisdiction. These type of FTA's would be officially recorded in court records but often result in no consequence to the defendant from the court.

To understand more about this perceived limitation of official court records, the pilot survey captured whether the pretrial defendant reported having an FTA in the last year that would be defined as "un-willful." One pretrial officer in the pilot feedback survey provided an example that exemplifies some of the potential issues of using court records in FTA tracking: "*FTA's in Colorado are very inconsistent. People get 'FTA'd' all of the time, even though they are actually in custody. Even if they show up late for court and the warrant was subsequently quashed, the FTA still shows up on the court record*." Furthermore, of the 1,731 pretrial defendants that responded to this inquiry during the CPAT-R interview, 43.39% (n = 751) reported having a willful FTA, 53.44% (n = 925) reported not knowing if they had an un-willful FTA, and 3% (n = 52) reported having an un-willful FTA in the last year. This stakeholder feedback was used to construct multiple FTA outcomes tested in the validation analysis. This measure defined FTA by the judicial response to the FTA– no, low, and high consequence.

Violent Arrest: Additional findings from the feedback survey respondents indicate perceived challenges to defining a prior violent arrest. Some respondents reported a high degree of difficulty requesting advising this risk item should be eliminated.

Time at current residence: Some respondents reported conceptual problems with the time at current residence risk factor. For example, one respondent noted that someone who recently purchased a home (within the last year) may still have strong ties to the community. This defendant would be scored the same number of points for this item as someone who may not have strong ties to the community. These two items were mentioned by numerous respondents and were subsequently dropped from the final recommended version of the tool as result of the bias analysis discussed below.

Part 1: Pilot CPAT-R Validation

The validation component of the phase three analysis sought to identify the assessment tool with the best estimated predictive performance, meaning the tool that best discriminates across risk scores and has the most consistent prediction of actual pretrial outcomes. To complete this analysis, the following research questions from the phase one analysis were revisited and revised:

1) How often are the CPAT-R, CPAT and CPATR-SV's risk level assignment consistent with pretrial outcome?

To answer this research question, within-individual comparisons were made across the three tools: the pilot CPAT-R, CPAT and CPATR-SV. The ROC curve and corresponding AUC score were estimated for each of the primary outcomes. The ROC and corresponding AUC score represent the ratio of true and false, positive and negative risk assessments. An AUC score of .50 or greater indicates that the tool is valid and assigns a risk score consistent with the outcome better than chance. AUC scores of existing validated pretrial risk assessment tools range from .50 - .75.[19] A test of equal AUC was then conducted to indicate if the observed differences in the AUC scores were systematic or rather, statistically significant.

Table 17. *Full sample – Released pretrial defendants AUC scores*

Outcome	Pilot CPAT-R AUC (BC CI)	CPAT – AUC (BC CI)	CPATR – SV (BC CI)	AUC diff, p-value
Primary				
New Arrest and/or FTA	.65* (.62-.68)	.60* (.57-.62)	.62* (.59-.65)	< .001
New Arrest	.59* (.56-.63)	.57* (.53-.60)	.56* (.52-.60)	< .001
FTA	.66* (.63-.69)	.60* (.57-.63)	.64* (.61-.67)	< .001
Secondary				
Violent	.57 (.44-.66)	.57 (.44-.67)	.55 (.43-.64)	.28
DV or order Violation	.52 (.42-.61)	.55 (.46-.64)	.51 (.40-.60)	.51
Serious	.60* (.55-.64)	.58* (.53-.62)	.57* (.52-.61)	<.001
Other	.60* (.53-.66)	.55 (.49-.62)	.55 (.49-.61)	<.001
FTA – No	.64* (.61-.67)	.60* (.56-.63)	.62* (.58-.65)	< .001
FTA – Low	.57 (.47-.66)	.55 (.47-.63)	.56 (.47-.66)	.86
FTA – High	.68* (.63-.72)	.58* (.53-.62)	.64* (.59-.69)	< .001

* = Estimate is likely beyond chance *n* = 1,486

a) Confidence Intervals (bootstrap 1000)

[19] 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 and implemented in correctional settings in the United States.* Lexington, KY: Council of State Governments. 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).

ROC curves were run for each assessment tool. Table 17 reports the estimates for the corresponding AUC scores and validation comparisons. Bootstrapping was applied to the ROC and corresponding AUC scores throughout the phase 3 analysis. This approach involves repeated sampling of the observed data approximating the sample distribution. This approach can produce sample estimates that are more likely to be observed in the larger population of pretrial defendants. Accompanying variable selection, bootstrapping can result in more predictive models.[20] All three of the tools validate according to the .50 significance threshold. This means that the AUC estimates were above this significance threshold and likely to be observed in the larger population of pretrial defendants. This can be visually confirmed with the ROC curve comparison plots for each tool diverging from the reference line (Appendix H). The CPAT-R demonstrated the greatest predictive performance across each of the three primary outcomes. The differences across the predictive performance of each tool was systematic ($p < .001$) indicating that these differences are also likely to be observed in the larger population of pretrial defendants. These estimates provide support that the CPAT-R maintains the best estimated predictive performance compared to the CPAT and CPATR-SV.

In addition to estimating the ability for each tool to quantify a discriminate risk level, calibration plots were also estimated. Calibration plots illustrate each tool's ability to assign a risk level that is consistent with the observed pretrial outcome. A well-calibrated tool will not only correctly assign an individual risk score that is relative to the risk of the population but will also assign a risk level that is consistent with the actual pretrial outcome. Calibration plots were estimated using STATA's pmcalplot command for each tool and primary outcome.

The LOWESS (Locally Weighted Scatterplot Smoothing) line on a calibration plot represents the average difference between what the risk score predicts (expected) and the actual outcome (observed). A divergence of the LOWESS line from the straight diagonal reference slope can indicate over- or under-fitting. This means that the tool either assesses more defendants as high risk than those who actually have failed pretrial outcome or it assesses more defendants as low risk than those who actually have a successful pretrial outcome. The calibration plots indicate a relatively good fit for each tool (Appendix H). For all of the tools, the LOWESS line is consistent with the reference line but diverges at higher levels of expected pretrial failure or rather, higher risk scores.

The ROC curve and corresponding AUC scores, as well as the difference in means tests were also estimated for the more specific new arrest type and FTA related outcomes defined by charge and judicial consequence. All three tools validated when assessing the risk of a serious new arrest, FTA with no court consequence and an FTA with a high consequence. Of these validated outcomes, the CPAT-R was estimated to systematically perform better ($p < .001$). Calibration plots were also estimated and indicate a good fit.

[20] Austin, P.C. & Tu, J.V. 2004. Bootstrap methods for developing predictive models. *The American Statistician, 58(2),* 131-137.

These findings indicate that of the pilot study tools, the CPAT-R maintains the best discriminate and predictive performance. These findings provide support to further assess the features of the CPAT-R and identify opportunities to make supervised improvements. To do so, the validation sub-sample is used to examine the tool's weighting and risk factors.

2) Are the CPAT-R's risk factor's weighted for the tool's best predictive performance?

To answer this research question, alternative approaches to weighing the pilot CPAT-R's risk factors were used and compared to each tool's predictive performance. The weighting approaches assessed in phase one were replicated and re-weighted using a dichotomous Burgess approach and margins weighting. Table 18 reports the dichotomous scoring that the Burgess weighting imposes. A Burgess weighted CPAT-R assigns 1 point to each risk factor that would sum to a score greater than 0. This approach gives each risk factor equal weight contribution to the sum of the risk score. The Burgess weighted CPAT-R ranges in score from 0 – 11. The margins reweighting approach weighs each risk factor according to its individual contribution to the likely change in odds of a pretrial outcome. The risk factors that were assigned a higher score result in a greater change in odds of new arrest and/or FTA.

Table 18. *CPAT-R re-weighted risk factors*

Risk Factor	Pilot CPAT-R Weight (points)	Burgess Weight (points)	Margins Weight (points)
Employment/Education	0/2	0/1	0/2
Time at current residence	0/1	0/1	0/4
Problems with alcohol or drugs	0/1	0/1	0/4
Prior Arrest	0/3	0/1	0/1
Arrest in the last year	0/3	0/1	0/6
Age at first arrest	01	0/1	0/3
Prior FTA	0/3	0/1	0/7
FTA within the last year	0/3	0/1	0/12
Pending charge at arrest	0/1/2	0/1	0/1
Active warrant	0/2	0/1	0/4
Prior violent arrest	0/1	0/1	0/5
Range	0 - 22	0 - 11	0 - 49

Table 19. *CPAT-R re-weighted comparisons*

Outcome	CPAT-R CV-AUC (BC CI)	CPAT-R Burgess CV-AUC (BC CI)	CPAT-R Margins CV-AUC (BC CI)
New Arrest/FTA	.6661* (.6275 - .6923)	.6628* (.6158 - .6830)	.6593* (.6164 - .6831)
New Arrest	.6333* (.5725 - .6615)	.6335* (.5668 - .6542)	.6287* (.5703 - .6609)
FTA	.6724* (.6217 - .6944)	.6657* (.6120 - .6842)	.6678* (.6224 - .6967)

* = Estimate is likely beyond chance *n* = 1,266

a) 10-fold cross-validated

b) Confidence Intervals (bootstrap 1000)

To assess the impact of this modification, k-fold cross-validation was employed. This technique compares the predictive performance of the pilot CPAT-R and reweighted versions. Cross-validation represents the average of multiple, randomly sampled AUC score estimates. It is useful for internal validation because it is more likely generalizable to the population.[21] This means that sample estimates using this approach are more likely to be observed in the larger population of pretrial defendants. The pilot CPAT-R cross-validated at a slightly higher AUC score across the combined new arrest and/or FTA and FTA outcomes. The pilot CPAT-R was estimated to have a slightly lower cross-validated AUC score for the new arrest outcome. This provides support that the 0-3 weighting approach may provide the optimal predictive performance compared to the re-weighted modifications.

3) Are the risk factors in the CPAT-R the best predictors of certain pretrial outcomes?

To further assess the risk factors included in the pilot CPAT-R for predictive performance, the tool's predictive performance was tested and compared to the predictive performance of the most predictive risk factors in the pilot CPAT-R. To do this, the variable selection process of LASSO regression was replicated. This procedure identified four predictive risk factors: arrest within the last year, prior FTA, FTA within the last year, and prior violent arrest.[22] K-fold cross validation was also conducted to compare this modification to the CPAT-R.

Table 20. *CPAT-R revised risk factor comparison*

Outcome	**CPAT-R CV-AUC (BC CI)**	**CPAT-R Risk Factor Revised CV-AUC (BC CI)**
New Arrest/FTA	.6661* (.6275 - .6923)	.6660* (.6311 - .6894)
New Arrest	.6333* (.5725 - .6615)	.6371* (.5848 - .6544)
FTA	.6724* (.6217 - .6944)	.6674* (.6244 - .6852)

* = Estimate is likely beyond chance *n* = 1,266

a) 10-fold cross-validated

b) Confidence Intervals (bootstrap 1000)

The CPAT-R was estimated to have a slightly higher cross-validated AUC score for the combined new arrest and/or FTA and FTA only outcomes compared to the revised risk item modification. The tool had a slightly lower estimated cross-validated AUC score for the new arrest only outcome. This finding provides support that the CPAT-R is a more comprehensive model than the revised risk factor modification.

[21] Rodriguez, J.D., Perez, A. & Lozano, J.A. (2009). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 32(3), 569-575.

[22] Risk factors: Arrest within the last year, β =.05; At least one prior FTA, β = .04; FTA within the last year, β = .16; prior violent arrest, β = .03

The validation estimates that were used to assess and compare the piloted tools' predictive performance indicated that the CPAT-R may discriminate risk and predict the observed outcome better than the CPAT and CPATR-SV. Further evaluation of the features of the CPAT-R, including its weighting and risk factors, indicated that the piloted CPAT-R without modification may also maintain the best predictive performance. These findings provided support that the unmodified, piloted CPAT-R was the best candidate for the accuracy equity and predictive parity analysis.

Part 2: Accuracy Equity and Predictive Parity Assessment

The second component of the phase three analysis involved comparing the best performing tool across sub-groups of pretrial defendants. The sub-groups examined include race/ethnicity, gender, and residential status. The goal of this assessment was to evaluate the accuracy equity and predictive parity of the tool and further modify the tool to reduce differences across the sub-groups. The validation sample was used to estimate these differences and inform modifications to the weighting and inclusion of certain risk factors. These modifications were internally cross-validated on the validation sample and subsequently tested on the test sub-sample.

Accuracy equity refers to the predictive performance of an assessment tool across sub-groups.[23] A pretrial risk assessment tool that has accuracy equity would assign risk scores at similar accuracy across sub-groups of pretrial defendants. This is estimated by comparing the ROC and corresponding AUC scores across sub-groups for systematic differences in predictive performance. Systematic refers to statistically significant differences that are likely in the population. Predictive parity was also assessed across sub-groups. Predictive parity refers to differences in assessment error as a result of sub-group membership.[24] Over-assessment error occurs when a tool assigns a high risk that does not result in an outcome consistent with a high risk assessment. Predictive parity was assessed using the validation sample's descriptive false positive and negative outcome and prediction error rates.

[23] Dieterich, W. Mendoza, C. & Brennan, T. (2016). COMPAS risk scales: Demonstrating accuracy equity and predictive parity. Northpointe Inc.
[24] Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153-163.

Table 21. *Validation sample – Released pretrial defendants sub-groups*

Sub-group	Average CPAT-R Score (*SD*)	Baserate for New Arrest and/or FTA (*SD*)	New Arrest/FTA AUC (BC CI)
Race/ethnicity[a]			
White (*n* = 656)	9.03 (5.45)	.31 (.46)	.70* (.65-.74)
Black (*n* = 186)	10.34 (4.81)	.30 (.46)	.60* (.52-.69)
Hispanic (*n* = 331)	10.13 (5.36)	.35 (.48)	.64* (.58-.70)
Sex[b]			
Male (*n* = 914)	9.83 (5.28)	.33 (.47)	.66* (.63-.70)
Female (*n* = 272)	9.99 (5.63)	.36 (.48)	.62* (.55 -.69)
Residential Status[c]			
Homeless (*n* = 206)	12.49 (4.85)	.43 (.49)	.62* (.55-.70)
Housed (*n* = 1,060)	9.16 (5.34)	.30 (.46)	.67* (.63-.70)

* = Estimate is likely beyond chance *n* = 1,266

a) $X^2 = 4.88$; β = -.16, BC CI: -.32 - -.02)

b) $X^2 = 1.27$; β = -.09, BC CI: -.43 - .19)

c) $X^2 = .90$; β = -.19, BC CI: -.51-.13)

d) Confidence Intervals (bootstrap 1000)

Descriptive statistics and performance measures were first estimated on the validation sample to identify differences across sub-groups. Table 21 reports the average CPAT-R risk score, baserates, and AUC scores for the primary outcomes. The baserates represent the rate of pretrial failure for each sub-group. The sub-groups were defined using information self-reported by the pretrial defendant during the assessment interview. Race/ethnicity was assessed across three categories: White/non-Hispanic, Black/non-Hispanic, and Hispanic/either White or non-White (subsequently referred to as White, Black, or Hispanic).[25] Gender was categorized as male or female and residential status was categorized as self-reported homeless or housed.

The average CPAT-R score varied about 1 point across the three race/ethnicity sub-groups (Table 21). Self-reported White pretrial defendants scored on average the lowest with an average risk score of 9.03 (*SD* = 5.45). The baserates for the combined new arrest and/or FTA outcome were comparable across race/ethnicity sub-groups spanning .04 total. Male and female defendant risk scores had a smaller mean difference with female pretrial defendants being assessed slightly greater and having a slightly higher baserate, on average. The greatest mean difference in risk score was across residential status. Self-reported homeless pretrial defendants have an average risk score of 12.49 (*SD* = 4.85) and a .13 difference in baserate of new arrest and/or FTA compared to self-reported housed pretrial defendants.

[25] Other categories of self-reported race/ethnicity category not included in analysis due to sample size limitations.

Estimates comparing the ROC and corresponding AUC scores across the sub-groups indicated potential differences in predictive performance but these differences were not statistically significant indicating they are not systematic. Table 21 reports AUC scores that range .10 across race/ethnicity, .04 across gender, and .05 across residential status. Chi-square estimates were not statistically significant, indicating that these observed differences are not systematic or rather, likely to be observed in the larger population of pretrial defendants. The sub-group coefficient for race/ethnicity does indicate potential differences in risk prediction across groups. This estimate was statistically significant indicating that the risk score could systematically vary according to group membership (β = -.16, BC CI: -.32 - -.02). These same systematic differences were not estimated across gender or residential status. These findings provide support for further examination of the differences in predictive parity across sub-groups to improve the tool.

Predictive parity in the validation sample was described across sub-groups for the tool, as well as the individual risk factors. To estimate the assessment error, the continuous risk score was split into high and low categories. To do this the cutpoint between risk category 2 and 3 was used. This threshold was considered the most appropriate because it approximates the risk category release decision.[26] This threshold was defined for each modified tool using regression classification regression trees (CART). This approach identifies meaningful differences in the probability of an outcome across a range of scores.

The false positive rate represents the proportion of pretrial defendants assessed as high risk (positive) who did not have a new arrest and/or FTA (false outcome). False negative rate represents those identified as low risk (negative) who did have a new arrest and/or FTA (false outcome). The failure and success prediction errors represent the rate of predictions for either anticipated outcome, pretrial success or failure that are made in error. The overall prediction error represents the total of both failure and success prediction error. Particular attention was paid to differences in the false positive rate across sub-groups. A sub-group with a higher false positive rate is proportionately more often assessed to be high risk but have a successful pretrial outcome. Applying this to the pretrial release decision, a false positive would occur if a pretrial defendant was not released on bond but would have had a successful pretrial outcome if they had been released on bond.

[26] Picard, S., Watkins, M, Rempel, M & Kerodal, A. 2019. Beyond the algorithm: Pretrial reform, risk assessment and racial fairness. *New York: Center for Court Innovation.*

Table 22. *Validation sample - Race/ethnicity assessment errors*

Modification	False Positive Rate			False Negative Rate			Failure Prediction Error			Success Prediction Error			Overall Prediction Error		
	W	B	H	W	B	H	W	B	H	W	B	H	W	B	H
Pilot CPAT-R[b]	.24	.37	.29	.52	.55	.48	.52	.66	.50	.23	.27	.27	.41	.50	.43
No prior violent arrest[c]	.33	.29	.29	.37	.51	.51	.51	.52	.49	.22	.26	.30	.34	.35	.37
No prior violent arrest and reduce weight prior FTA[d]	.31	.41	.40	.39	.45	.38	.53	.63	.54	.20	.25	.26	.34	.42	.40
No prior violent and prior FTA[e]	.17	.21	.22	.64	.75	.67	.50	.66	.55	.25	.29	.32	.31	.37	.38

a) White n = 656, new arrest/ FTA baserate = .31; n = 1,173
Black n = 186, new arrest/FTA baserate = .30; 1,173;
Hispanic n = 331, new arrest/FTA baserate = .35.
b) The optimal high/low threshold is risk score 12, β = .42, p< .001.
c) The optimal high/low threshold is risk score 9, β = .45, p< .001.
d) The optimal high/low threshold is risk score 9, β = .45, p< .001.
e) The optimal high/low threshold is risk score 10, β = .47, p< .001.

Table 22 reports the assessment errors across race/ethnicity for the full model across self-reported race/ethnicity for four different modifications to the pilot CPAT-R. A difference in false positive rates across sub-groups for the pilot CPAT-R was .13. The assessment errors were estimated for each risk factor. The high/low risk threshold was defined using the risk factor score (0 = low risk and 0 > high risk) (Appendix I). The greatest difference across sub-groups was identified for prior violent arrest (difference = .33) and having at least 1 prior FTA (difference = .23). Differences across overall prediction errors were .15 for both variables across sub-groups.

The tool modifications in Table 22 include removing the prior violent arrest risk factor, reducing the prior FTA risk factor weighting from 3 to 2 points, and removing both prior violent arrest and prior FTA together. A comparable balance in false positive rate was accomplished when removing prior violent arrest or both the prior violent arrest and prior FTA risk factors. Removing both risk factors did result in a meaningful increase in false negative rates for all sub-groups. This provides support that the tool modification removing only the prior violent risk factor is the best candidate to assess errors across gender.

Table 23. *Validation sample - Gender assessment errors*

Modification	False Positive Rate		False Negative Rate		Failure Prediction Error		Success Prediction Error		Overall Prediction Error	
	M	F	M	F	M	F	M	F	M	F
No prior violent arrest[b]	.35	.39	.41	.39	.55	.53	.24	.26	.37	.39

a) Male, *n* = 914, new arrest/FTA baserate = .33; Female, *n* = 272, new arrest/FTA baserate = .36 *n* = 1,186
b) The optimal high/low risk threshold is risk score 9, β = .45, p< .001.

Table 23 reports the assessment errors across gender for the modified tool with the prior violent arrest risk factor removed. Female defendants were estimated to have a slightly higher false positive rate (.39) compared to males (.35) with a difference of .04. This accompanied with comparable prediction error across sub-groups provides support to move forward with the modified CPAT-R that has the prior violent arrest risk factor removed to assess across residential status.

Table 24. *Validation sample - Residential status assessment errors*

Modification	False Positive Rate		False Negative Rate		Failure Prediction Error		Success Prediction Error		Overall Prediction Error	
	H	NH	H	NH	H	NH	H	NH	H	NH
No prior violent arrest[b]	.60	.31	.26	.45	.51	.57	.33	.22	.45	.35
No prior violent arrest and time at residence2[c]	.46	.26	.40	.53	.50	.56	.36	.24	.44	.34
No prior violent arrest, time at residence and reduce weight employment/education[d]	.47	.28	.37	.50	.49	.47	.34	.23	.42	.35
No prior violent arrest, time at residence, and employment education[e]	.32	.21	.51	.63	.46	.57	.36	.26	.40	.34
No prior violent arrest, time at residence and reduce weight active warrant[f]	.50	.29	.38	.49	.51	.57	.37	.23	.45	.35
No prior violent arrest, time at residence and active warrant[g]	.29	.16	.58	.73	.48	.57	.39	.27	.42	.33

H = Homeless, NH = Housed *n* = 1,266
a) Homeless, *n* = 206, new arrest/FTA baserate = .43; Housed, *n* = 1,060, new Arrest/FTA baserate = .30
b) The optimal high/low threshold is risk score 9, β = .45, p <.001
c) The optimal high/low threshold is risk score 11, β = .42, p< .001.
d) The optimal high/low threshold is risk score 10, β = .43, p<.001.
e) The optimal high/low threshold is risk score 11, β = .44, p < .001.
f) The optimal high/low threshold is risk score 10, β = .40, p <.001.
g) The optimal high/low threshold is risk score 12, β = .39, p<.001

Table 24 reports the assessment errors across residential status, self-reported homeless and housed. The modified tool that had the prior violent arrest risk factor removed indicated a difference of .29 in false positive rates across self-reported residential status and .10 difference in the overall prediction error. The individual risk factor assessment errors indicate substantial differences in false positive for the time at current residence (.75), as well as the employment or education status (.40) (Appendix I). The greatest reduction across the sample's false positive rate was when prior violent arrest and time at current residence were removed or when prior violent arrest, time at current residence and employment status were removed. Removing all three risk factors was estimated to result in a greater difference across the false negative rate, providing support that removing the prior violent arrest and time at current residence risk factors produces a balanced tool modification.

The assessment errors were estimated on the validation sample for this tool modification across the other sub-groups, race/ethnicity and gender. Table 25 illustrates that across each sub-group the descriptive estimates indicate there were relative differences in the estimated errors, but the prediction errors across sub-groups differences remain comparable for this modification.

Table 25. *Validation sample - Balance candidate assessment errors across sub-groups*

Sub-group	False Positive Rate	False Negative Rate	Failure Prediction Error	Success Prediction Error	Overall Prediction Error
Race/Ethnicity[a]					
White	.26	.51	.54	.26	.34
Black	.34	.55	.64	.26	.40
Hispanic	.31	.49	.52	.28	.37
Sex[b]					
Male	.28	.50	.54	.25	.35
Female	.35	.52	.56	.31	.41
Residential Status[c]					
Homeless	.46	.40	.50	.36	.44
Housed	.26	.53	.56	.24	.34

a) White = 656, new arrest/ FTA baserate = .31; Black = 186, new arrest/FTA baserate = .30; Hispanic = 331, new arrest/FTA baserate = .35 *n* = 1,266

b) Male: *n* = 914, new arrest/FTA baserate = .33; Female: n = 272, new arrest/FTA baserates male = .36

c) Homeless: *n* = 206, new arrest/FTA baserate = .43; Housed: *n* = 1,060, new arrest/FTA baserate = .30

Finally, the relationship between the tool's risk score and new arrest and/or FTA outcome were compared. To do this, statistical interactions between sub-group membership and risk score were estimated. These estimates can inform whether the modified tool, balance candidate meaningfully varies as a result of sub-group membership. Said another way, whether a pretrial defendant's risk score partially depends on their race/ethnicity, sex, or residential status. Table 26 reports the model estimates. Interaction terms for each sub-group were not statistically significant, indicating that the predictive performance of the modified tool is not systematically related to the tool's risk assignment.

Table 26. *Validation sample - Balance candidate interaction with sub-group membership*

Sub-group	***OR***	***SE***
White[a]		
CPAT-R_Score	1.10***	.02
White (0/1)	.64	.18
Score * White	1.05	.03
Black[b]		
CPAT-R_Score	1.10***	.02
Black (0/1)	.85	.40
Score * Blk	.99	.04
Hispanic[c]		
CPAT-R_Score	1.09	.05
Hispanic (0/1)	1.06	.65
Score * Hispanic	1.01	.05
Male[d]		
CPAT-R_Score	1.08*	.03
Male (0/1)	.60	.19
Score * Male	1.04	.03
Homeless		
CPAT-R_Score	1.12***	.02
Homeless (0/1)	1.60	.68
Score * Homeless	.99	.04

a) $X^2 = 91.41$, $n = 1{,}266$ $p < .05$*, $p < .001$***
b) $X^2 = 27.48$, $n = 610$
c) $X^2 = 20.68$, $n = 424$
d) $X^2 = 76.60$, $n = 1{,}186$
e) $X^2 = 92.48$, $n = 1{,}266$

Assessment errors were also calculated for the CPAT and CPATR-SV (Appendix J). These estimates indicate no meaningful improvement in balance across sub-groups for the validation sample. This analysis provides support that the best performing and balanced tool modification is the piloted CPAT-R with the prior violent arrest and time living at current residence risk factors removed. The next section details a supplemental analysis of the reliability and features of this modified tool.

Part 3: Reliability and Feature Analyses

The reliability and feature analyses assessed the agreement between official records and self-reported information, and the features of the risk factor definitions. The features that were assessed were informed by feedback from stakeholders, observation of assessment training, observation of the interview, and assessment processes in the participating counties.

Self-report Reliability

To assess the consistency across information sources, the agreement between self-report and official criminal/court history measures were compared. This analysis informed whether pretrial defendants provided meaningful different information about their criminal and court history compared to official records.

Table 27. *Self-report and official reliability*

Risk factor	**Agreement %**	**Kappa (SE)**
Prior Arrest	67.71	-.04 (.63)
Arrest in the last year	70.40	.41 (.02)***
Prior FTA	58.45	.16 (.02)***
FTA in the last year	63.85	.22 (.02)***
Pending charge at arrest	71.93	.30 (.02)***
Active warrant	68.36	.29 (.02)***

*p<.05 *** p < .001

Table 27 reports kappa coefficients across self-report and official record confirmed assessment scores. This estimate represents the inter-rater reliability and has a range of -1 to +1. A coefficient closer to +1 indicates greater agreement between the measures. The kappa estimates indicate low to moderate agreement for all risk factors that can be confirmed with official records except prior arrest (2+). About 95% of defendants reported having at least two prior arrests and 73% of the assessment responses confirmed with criminal history records indicated the same. This provides support when self-reported criminal history does not agree with official records that defendants may over-report their criminal history.

Employment/Education: Of those who reported being employed and assessed by the CPAT-R, 81% (n = 871) provided self-reported employed status. Of those self-reporting employment, 90.24% (n = 786) reported full-time employed. The remaining 9.76% self-reported being a student, receiving disability, being a stay-at-home parent or retired. Differences across the average new arrest and/or FTA outcome were assessed across self-reported employed and other employment (t = -.003, p =.99, n = 860). No meaningful differences in pretrial outcomes across employment type were estimated. This estimate supports including all of the self-reported types of employment in this risk factor definition.

Current or Past Alcohol or Drug Problem: The difference in the predictive performance of the self-reported alcohol or drug problem was assessed across current and past problem. Pretrial defendants were asked if their self-reported alcohol or drug problem was in the past, is current or both. A modified version of this risk factor was recoded to score 1 point if the defendant reported having a current problem. This resulted in 430 defendants no longer being scored as a 1 for this risk item because they only reported having an alcohol or drug problem in the past. A modified version of the recommended tool was cross-validated using the validation sample and was found to maintain predictive performance (CV-AUC = .66, BC CI: 62-.68).[27] This provides support that removing the past only feature of this risk item definition may not interfere in the predictive performance of the tool.

Age at first arrest: Pretrial defendants interviewed for the CPAT-R were asked to self-report their age at first arrest. This measure was compared to an age at first arrest confirmed with criminal history records by the pretrial officer. The average self-reported age at first arrest was 20 years of age (*SD* = 7.80) and official record confirmed age at first arrest was 22 years of age (*SD* = 8.42). A difference in means test indicates this difference is not systematic (*t* = -1.03, *SE* = .20, *n* = 680, p = .30) and supports no further modification to this measure.

Confirming information: This study pilot tested confirming information in several of the participating counties. Of the 792 recorded attempts made to successfully confirm information with a reference, 7.20% were identified as successful. Coupled with the feedback survey conducted after the CPAT-R pilot, this process was determined to be a time-consuming effort without much benefit to the pretrial investigation process. More research needs to be done on the utility and benefit of this implementation feature, but at this time the researchers have not identified a strong support for this practice for the purposed of administering pretrial risk assessment. Findings from phase 1 and 2 do indicate that this practice may still have value for effectively carrying out the pretrial process.

These estimates were used to inform further modifications to the tool including removing the past alcohol or drug feature of this self-reported risk factor. This modified tool was tested using the test sub-sample for predictive performance, accuracy equity and predictive parity. The modified tool tested was detailed in Table 28. Risk categories were defined using the regression classification tree (CART) analysis. This approach identifies the greatest change in the probability of a new arrest and/or FTA outcome across the range of risk scores.

[27] k-fold cross-validation, k = 10

Table 28. *Modified tool*

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 yoa or older, 1 = 20 yoa 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) Risk categories: Category 1, scores 0-7 ($\beta = .25$, $p<.001$); Category 2, scores 8 – 11 ($\beta = .41$, $p<.001$); Category 3,scores 12 – 14 ($\beta = .34$, $P < .001$), Category 4, scores 15 – 20.

Test Analysis

Table 29 reported the predictive performance estimates for the modified tool for both the primary and secondary outcomes using the test sub-sample. These estimates are used to inform recommendation for the CPAT-R. The predictive performance of the modified tool validates well for the primary outcomes. For the combined new arrest and/or FTA outcome the tool the AUC score is estimated to be above the .50 threshold (AUC = .66, BC CI: .61-.71). This means that the tool is likely to estimate actual pretrial outcome better than chance, and falls into the "good" predictive validity range by Desmarais & Singh (2013).[28] Review of the ROC curve and corresponding calibration plots also indicate good calibration (Appendix K). The modified tool validly predicts FTA outcome, no and high consequence but does not validate on the other type of new arrest or FTA secondary outcomes.

Table 29. *Test sample –AUC scores and confidence intervals of primary and secondary outcomes*

Outcome	AUC (BC CI)
Primary Outcome	
New Arrest and/or FTA	.66* (.61-.71)
New Arrest	.57* (.51-.64)
FTA	.69* (.64-.74)
Secondary Outcome	
Violent	.54 (.27-.76)
DV or order Violation	.57 (.40-.74)
Serious	.57 (.48-.64)
Other	.57 (.48-.66)
FTA – No	.64* (.57-.69)
FTA – Low	.57 (.44-.70)
FTA - High	.73* (.66-.78)

* = Estimates likely beyond chance $n = 500$

a) Confidence Intervals (bootstrap 1000)

Accuracy equity was assessed across sub-groups using the test sub-sample. No systematic differences across AUC scores are estimated across race/ethnicity ($X^2 = 1.42$), gender ($X^2 = .23$), and residential status ($X^2 = -.95$) (Table 30). The systematic difference estimated across race/ethnicity for the piloted CPAT-R was no longer statistically significant when estimated for the modified tool ($\beta = -.05$, BC CI: -.34 - .21).

[28] Desmarais, S. L., and Singh, J P. (2013). *Risk assessment instruments validated and implemented in correctional settings in the United States.* Lexington, Kentucky: Council of State Governments.

Table 30. *Test Sample - AUC score for modified tool across sub-groups outcome new arrest/FTA*

Sub-group	New Arrest/FTA AUC (CI)
Race/ethnicity[a]	
White	.71* (.65-.78)
Black	.62* (.50-.76)
Hispanic	.70* (.62-.79)
Sex[b]	
Male	.69* (.63-.74)
Female	.66* (.56-.76)
Residential status[c]	
Homeless	.69* (.63-.74)
Not Homeless	.62* (.50-.74)

*= Estimates likely beyond chance $n = 500$

a). X^2= 1.40; β = –.05 (BC CI: -.34-.21)

b) X^2 = .23; β = -.14 (BC CI -.73-.36)

c) X^2 = -.95; β = .18 (BC CI .79-.63)

Predictive parity was also estimated for the test sub-sample. Table 31 reports the descriptive rate and predictive errors. The differences in rate and prediction error are comparable to the validation sub-sample estimates.

Table 31. *Test sample – Modified tool assessment errors*

	False Positive Rate	False Negative Rate	Failure Prediction Error	Success Prediction Error	Overall Prediction Error
Race/Ethnicity[a]					
White	.19	.63	.48	.30	.35
Black	.13	.66	.41	.29	.32
Hispanic	.13	.60	.34	.30	.30
Sex[b]					
Male	.17	.68	.48	.33	.36
Female	.22	.46	.36	.30	.32
Residential Status[c]					
Homeless	.36	.48	.41	.43	.42
Housed	.15	.57	.47	.29	.43

a) White, non-Hispanic = 249, new arrest/FTA baserate = .36; Black = 82, new arrest/FTA baserate = .35; Hispanic = 128, new arrest/FTA baserate = .38

b) Male = 345, new arrest/FTA baserate = .37; Female = 124, new arrest/FTA baserate = .42

c) Homeless = 88, new arrest/FTA baserate = .50; Housed = 412, new arrest/FTA baserate = .34

The test analysis estimates conducted above provide support for the use and adaptation of the modified tool. The following section of the report includes final recommendations informed by the analyses in all three phases of the project.

Recommendations

The overall goal of this project was to create the strongest and most equitable version of the CPAT that could be implemented in counties across Colorado. This was accomplished through over the three phases of the study and employed feedback from criminal justice stakeholders, access to criminal justice data from multiple sources throughout the state, and the utilization of the most current analytical procedures used in risk assessment, the following recommendations are made.

Use of the Recommended CPAT-R: The recommended version of the CPAT-R was estimated to have the best predictive performance and the most balance across accuracy equity and predictive parity. Below is the recommended CPAT-R and full sample estimates of the rate and odds of pretrial success for new arrest and FTA outcomes. Success is defined as a new arrest or FTA not occurring.

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

Risk Category	Rate of success – new arrest[a]	Odds of success – new arrest[b]	Rate of success – FTA[c]	Odds of success – FTA[d]
1	97%	8.48 greater	91%	10.01 greater
2	82%	4.58 greater	71%	2.36 greater
3	78%	3.50 greater	67%	2.02 greater
4	74%	3.24 greater	66%	1.97 greater

a) The proportion of those in the risk category that are not arrested during the pretrial release period. n = 1,858

b) The odds of no new arrest vs. a new arrest occurring (e.g. Category 1 defendants have an 8.48 greater odds of no new arrest compared to a new arrest.)

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

d) The odds of no FTA vs. an FTA occurring (e.g. Category 1 defendants have a 10.01 greater odds of no FTA compared to an FTA occurring

Accuracy and Balance: Findings provide support that the recommended CPAT-R has greater predictive performance at assigning a pretrial risk score compared to the CPAT and CPATR-SV. It has also been assessed for differences across sub-groups. Robust evaluation methods were used to estimate predictive performance and informative, descriptive estimates were used to assess differences across sub-groups. This provides strong support for the current recommendations, as well as for on-going comprehensive evaluation of the recommended CPAT-R to ensure that test estimates remain consistent over time.

Comprehensive pretrial risk assessment evaluation should include assessment of both the predictive performance and differences across sub-groups in the tool's performance. It should also include evaluation of the fidelity to the tool's implementation protocol. Findings from phases one and two indicate that implementation of the risk assessment protocol could vary across jurisdiction. Monitoring fidelity to the tool's implementation protocol allows for divergence from the implementation protocol to be identified and assessed for its impact on the tool and release decision. The tool's implementation protocol is defined in the implementation manual that accompanies this report.

Appropriate and Standardized Use: The recommended CPAT-R is constructed to inform pretrial release decisions in conjunction with professional discretion. Findings from phases one and two indicate that assessment scores inform release decisions along with a variety of other factors. It is recommended that decision-makers continue this practice and incorporate the predictive performance and relative differences in assessment errors across sub-groups with particular attention to the false positive rate. Furthermore, findings indicate that pretrial risk assessment tools are used to inform decisions about release and supervision. The participating counties used decision-making matrices along with assessment and other relevant information about pretrial defendants to make decisions about bond conditions. This study was limited in testing the pretrial release and bond condition decisions informed by the recommended CPAT-R. Evaluation of the role of the recommended CPAT-R in the bond condition decision could provide valuable insight into effective uses of the tool.

The recommended CPAT-R validated for the primary outcomes new arrest and/or FTA, new arrest and FTA. Predictive performance test estimates across secondary outcomes defined by type of new arrest and degree of FTA did not consistently validate. Furthermore, these outcomes were not assessed for accuracy equity or predictive parity. It is recommended to use the tool's risk score to inform decisions that weigh only the primary outcomes. Assessment informed decisions that pertain to the secondary outcomes should be made using other risk assessment tools validated for pretrial, offense-specific outcomes.

Findings relevant to the standardization of use indicate that those involved in carrying out the pretrial process highly value education about pretrial risk assessment tools, as well as training about how to implement a risk assessment tool. Interest in receiving more education and training was reported by pretrial officers and stakeholders across role and jurisdiction. It is recommended that the audience for formal education and training opportunities expand to also include judges, prosecutors, and public defenders. These efforts can ensure standardization of administration of the tool by pretrial officers, as well as standardization of interpretation for those that use pretrial risk assessment to inform bond arguments and release decisions. This recommendation could be implemented through re-scaling or re-formatting (i.e., online) current education and trainings efforts.

APPENDIX A

Phase 1 Perception Survey Items

Q1. What is your role in the criminal justice system (Select which most closely corresponds with your role):
o Pretrial Services (1)
o Judge (2)
o Defense (3)
o Prosecution (4)

Q2 How do you feel about the CPAT in its current state?

_____

Familiarity How Familiar are you with the CPAT? Scale = 1-10, with 1 = not familiar at all, 5= understand the instrument conceptually, 10 = very familiar, know all of the items.

0 = Not familiar at all 5 = Understand the instrument conceptually 10 =Very familiar, know all of the items

0 1 2 3 4 5 6 7 8 9 10

Q3. What do you like about the CPAT?

_____

Q4. What do you feel could be improved with the instrument?

_____

Q5. What do you believe to be the benefits of the tool?

_____

Q6. As part of your job duties, do you conduct CPAT interviews?
o Yes (1)
o No (2)

Skip To: Q16 If As part of your job duties, do you conduct CPAT interviews? = No

Q7. Please describe the steps you go through when using the CPAT

_____

Q8. How many CPAT interviews do you conduct in one week?

_____

Q9. Do you confirm the information solicited in the CPAT interview with outside sources?
o Yes (1)
o No (2)

Q10. What (if any) additional resources or training do you think would be beneficial for working with the CPAT?

_____

Q11. How much time do you have to complete each assessment?

_____

Q12. Are your pretial risk assessment reports reviewed by a second party before being submitted to the court?

o Yes (1)
o No (2)

Q13. If Yes, by who reviews your pretrial report?

_____

Q14. Rank the CPAT tasks you find most important to least important. (important = most valuable in achieving a reliable CPAT score, reliable info etc.)

_____ Interview (1)
_____ Confirming information (2)
_____ Criminal history checks (3)
_____ victim interviews (4)
_____ CPAT assessment (5)

Q15. What are your thoughts on the interview portion of the CPAT overall?

_____

Q16. Are there items on the CPAT that you feel like should be eliminated? If so, which items and why?

_____

Q17. Are there items on the CPAT that you feel should be added? If so, why these items?

_____

Q18. Do you think confirming CPAT interview information with outside sources is important? Why or Why not?

_____

Q19. Would you support a risk assessment tool without an interview?

o Yes (1)
o No (2)

Q20. Why or Why Not?

_____

Q21. How do you think judges feel about the CPAT?

_____

Q22. In what ways do you think we could increase judicial buy-in? (Buy-in = use of CPAT, adhering to recommendations consistent with CPAT score)

_____

Q23. How do you think your District Attorney feels about the CPAT?

_____

Q24. Are there ways you think we could increase DA support of this instrument?

_____

Q25. How do you think your Public Defender feel about the CPAT?

_____

Q26. Are there ways you think we could increase Public Defender support of this instrument?

_____

Q27. Are there any definitions of terms used for the CPAT that you feel are unclear? If so, please describe,

_____

Q28. Are there Pre-trial definitions that should be standardized throughout the different counties?

_____

Q29. Overall, what do you think about risk assessment tools in the criminal justice system?

_____

Q30. How valuable are risk assessment tools to your job?

Not Valued at All Highly Valuable

0 1 2 3 4 5 6 7 8 9 10

Start of Block: Demographics

Q31. What County do you currently work in?

_____

Q32. How long have you worked in your current position? (in years)

_____

Q33. How long have you worked in the criminal justice field? (in years)

_____

Q34. What is your current position?

_____

Q35. What is your current age? (in years)

Q36. What is the highest degree you have received?
- o High School Diploma (1)
- o Some College (2)
- o Associate's Degree (3)
- o Bachelor's Degree (4)
- o Master's Degree (5)
- o Juris Doctor (6)
- o Other (7) _____

Q37. What is your gender?
- o Male (1)
- o Female (2)
- o Other (3)

Q38. Do you have any additional feedback?

_____

APPENDIX B

Phase 2 Focus Group Questions

Focus Group Questions: Stakeholder Groups

1. How do you feel about the CPAT in its current state?
   a. What do you like about it?
   b. What could be improved?

2. What is the utility of the CPAT to you?
   a. How do you use it?
   b. In practice, what do you do with the CPAT score?
   c. When should the CPAT scores be overridden?
      i. When do you override them?

3. What do you think could be done to help buy-in within your agency?

4. How many of you have been trained to use the CPAT?
   a. By whom?
   b. Would you support more frequent trainings?

5. If a tool was developed without an interview, would you be in support?
   a. What is the value of the interview to you?

6. Modifications:
   a. Specific questions
   b. Scoring Groups
   c. Offenses
   d. Would you support a tool with two risk groups?
      i. Risk to reoffend, risk to FTA

7. What type of training/resources would be useful for you?

8. Are there any items that need to be reconsidered?

9. What items (if any) could be added?

Focus Group Questions: Pretrial Officers

1. How do you feel about the CPAT in its current state?
    a. What do you like about it?
    b. What could be improved?
        i. Terms/definitions that are unclear?

2. What is the utility of the CPAT to you?
    a. Are there components of the CPAT process that are more time consuming than others?
    b. When do you believe CPAT scores be overridden?
        i. When, if at all, do you override them?

3. What aspects of the interview do you believe make the CPAT effective?
    a. What is the value of the interview to you?
    b. If a tool was developed without an interview, would you be in support?

4. How many of you have been trained to use the CPAT?
    a. By whom?
    b. Would you support more frequent trainings?
    c. What type of training/resources would be useful for you?

5. Does you agency have someone review your CPAT reports before they are submitted?
    a. Who typically does this? How important do you think this step is?

6. Does agency confirm CPAT interview information?
    a. If Yes
        i. With who?
        ii. What challenges do you confront when confirming information?
    b. If No:
        i. Why not?
        ii. Are there any methods you use to assure the information is accurate?

7. What do you think could be done to help buy-in of the tool?

8. Modifications:
    a. Specific questions
    b. Scoring Groups
    c. Offenses
    d. Would you support a tool with two risk groups?

        Risk to reoffend, risk to FTA

9. Are there any items that need to be reconsidered?

10. What items (if any) could be added?

APPENDIX C

CPAT-R Variable Selection List

CPAT risk factors: having a home or cell phone, owning or renting one's residence, contributing to residential payments, past or current problems with alcohol, past or current mental health treatment, age at first arrest, past jail sentence, past prison sentence, having active warrants, having pending cases, currently on supervision, history of revoked bond or supervision

Criminal history: prior violent arrest, prior domestic violence arrest, age at first arrest (continuous), prior arrest recency (5-yrs, 1-yr, 6-mos, 1-mon), prior arrest frequency (number, number in 5-yrs, number in 1-yr, number in 6-mos, number in 1-mon

Commuinity stability: time at residence, employment status

Offense type/severity: Current charge violent, felony charge, pending charge at arrest (felony, misdemeanor)

Court appearance history: Prior FTA (number, number in 5-yrs, number in 1-yr, number in 6-mos, number in 1-mon

APPENDIX D

Phase 1 CPAT-R Internal Validation Estimates

Pretrial Outcome	**CPAT: AUC (Confidence Interval)**	**CPAT-R: AUC (Confidence Interval)**
New Arrest or FTA – at all	.58* (.56 - .60)	.72* (.69 - .75)
FTA – at all	.54* (.52 - .56)	.70* (.65 - .75)
FTA – no consequence	.53* (.49 - .57)	.71* (.61 - .80)
FTA – low consequence	.53* (.49 - .56)	.63* (.51 - .75)
FTA – high consequence	.54* (.51 - .56)	.70* (.64 - .76)
New Arrest – at all	.54* (.52 - .57)	.72* (.69 - .75)
New Arrest – violent	.54 (.45 - .63)	.62* (.50 - .74)
New Arrest –DV or order violation	.51 (.45 - .57)	.52 (.43 - .60)
New Arrest – serious	.58* (.54 - .61)	.61* (.56 - .66)
New Arrest – other	.54* (.52 - .57)	.62* (.58 - .66)

* = Estimate is likely beyond chance

a) Estimates based on reduced sample, *n* = 3,124

b) Confidence Intervals (bootstrap 1000)

APPENDIX E

CPAT-R Pilot Survey

Date _____ Interviewing Officer Initials: _____

Name: _____ FBI#: _____SID#: _____

DOB: _____ Pretrial ID: _____

Sex: M / F / Other

Race: (Circle all that apply): American Indian or Alaska Native / Asian /

Black or African American / Native Hawaiian or Other Pacific Islander /

White / Other: _____ / Refused

Ethnicity: Hispanic or Latino / Non-Hispanic or Latino

Address: _____

Case #: _____

Current Charge(s): _____

_____

Confirming Contact Information:

Name: _____

Relationship: _____

Phone Number: _____

Confirming Information:

Were you able to make contact: Yes / No

Was the information confirmed: Yes / No

If no, what information was NOT confirmed (Circle all that apply):

Home/Cell Phone / Rent or Own / Contributing to Rent /

Employment or Education / Time at residence

If no interview was conducted, please explain why: _____

1. Employment/education

____ (0) Currently employed (full or part-time)/student /disability/retired/ stay-at-home parent or

____ (2) Not currently any of the above

Please select which applies:

____ Employed ____ Student

____ Disability ____ Retired

____ Stay-at-home parent/guardian

---

2. Time at current residence

____ (0) One year or more

____ (1) Less than one year

Time at residence:

____ Month(s) ____ Year(s)

Less than one month: ____ days

____ No residence/homeless

---

3. Problems with alcohol or drugs

____ (1) Yes

____ (0) No

If yes, which applies:

____ Alcohol ____ Drugs ____ Both

Self-reported identified problem:

____ Current ____ Past

---

4. Number of prior arrests

____ (3) At least two

____ (0) Less than two

Self-reported: Yes / No / Unknown

---

5. Age at first arrest

____ (1) 20 or younger

____ (0) 21 or older

Self-reported age: ___ / Unknown

Criminal history age: ___

---

6. At least 1 arrest in past year?

____ (3) Yes

____ (0) No

Self-reported: Yes / No / Unknown

---

7. Prior FTA

____ (3) Has at least 1 prior FTA

____ (0) Never had an FTA

Self-reported: Yes / No / Unknown

---

8. FTA within the last year

____ (3) Has FTA in last year

____ (0) No FTAs in last year

Self-reported: Yes / No / Unknown

Any un-willful FTAs: Yes / No

9. Pending charge at time of arrest

____ (0) None

____ (1) Misdemeanors only

Self-reported: Yes / No / Unknown

____ (2) At least one felony

---

10. Active warrant at time of arrest
____ (2) Yes
____ (0) No

Self-reported: Yes / No / Unknown

---

11. Prior violent arrest
____ (1) Yes
____ (0) No

If yes, please list all charge(s) for prior violent arrest:
_____
_____
_____
_____

APPENDIX F

CPAT-R Pilot Study Implementation Guide

Colorado Pretrial Assessment Tool (CPAT) and CPAT-Revised Pilot Study Implementation Guide

Victoria A. Terranova, PhD

Kyle C. Ward, PhD

Department of Criminology and Criminal Justice

University of Northern Colorado

The primary objective of this pilot study is to test the predictive accuracy of the Colorado Pretrial Assessment Tool (CPAT), as compared to modified versions of the CPAT. This guide and associated tools were developed through data collected from seven pretrial service agencies throughout Colorado, as well as observations and focus groups with the participating counties. While staffing, timing, and other resources may differ across jurisdictions, it is understood that all of these steps may be difficult to complete. This guide provides instruction on standard practice for counties participating in the pilot study to strive toward.

The modified versions of the CPAT is the CPAT-Revised (CPAT-R). The CPAT-R is a modified version of the CPAT that is designed to assign risk of both failure to appear (FTA) and re-arrest during the pretrial period. In order to collect data on the criminal history items on the CPAT-R, we ask that if a defendant is not available/refuses an interview, that pretrial officers fill out items 4-11, as well as several clarifying questions on the CPAT-R using criminal history records. Instructions for the pilot study protocol are detailed below.

During the pilot study interviewing officers will implement one of two risk assessment treatments per pretrial defendant: A) CPAT interview only or B) both CPAT-R and CPAT interviews. If a pretrial defendant is not available for an interview, the CPAT-R should still be utilized, but the interview questions (1-3) can be omitted. This pilot is estimated to span from November 13, 2018 to February 13, 2019. The risk assessment to be used at each case will be randomly assigned. Each participating site will be provided a pilot study CPAT assignment list. This list will inform pretrial officers whether they will administer the CPAT-R in addition to the original CPAT. If CPAT-R is to be administered, first administer the CPAT-R followed by the CPAT. If the defendant is unavailable for an interview, regardless of treatment assignment, fill out the interview-less items of the CPAT-R (i.e., items 4-11), as well as several of the clarifying questions detailed below.

Included in this instructional guide are:

-Interview and investigation steps for both interview treatments,

-Definitions for the CPAT-R risk items

-Recommended question wording, and

-Example information confirmation questions.

Please direct further questions to your designated CPAT-R pilot study representative. We appreciate your time and assistance with the pilot study.

Implementation Guide

A) CPAT interview only steps:

1. Identify from daily docket which clients will receive a CPAT interview.
    a. This will differ based on each jurisdiction's capacity and statute.

2. Conduct criminal history checks prior to the interview.
    a. Use court records and cross reference with CCIC/NCIC to determine age of first arrest, past jail or prison sentence, if currently on supervision, prior FTAs, and history of revocation.

3. Conduct CPAT Interview with client
    a. This could be conducted with the use of a computer, in person with a paper form, or with video service depending on the capacity of your jurisdiction.
    b. During the interview, obtain the phone number of a reference who can confirm the CPAT information. This could be a family member, spouse, friend, or roommate. If the defendant identifies the alleged victim, ask for another reference. Do not contact the alleged victim as a reference

4. Score CPAT
    a. Call the reference to confirm information from items 1-3:

        1: Having a home or cell phone
        2: Owning or renting one's residence
        3: Contributing to residential payments
        *Example confirming questions are included on the last page of this document.

    b. Rely on self-report data for items 4 and 5:

        4: Past or current problems with alcohol
        5: Past or current mental health treatment

    c. Use criminal history information from CCIC/NCIC, and Colorado court records to confirm items 6-12:

        6: Age at first arrest
        7: Past jail sentence
        8: Past prison sentence
        9: Having active warrants
        10: Having pending cases
        11: Currently on supervision
        12: History of revoked bond or supervision

5. Use the **CPAT** score to inform pretrial release recommendation per department protocol.

B) Both CPAT-R and CPAT interview steps:

1. Identify from daily docket which clients will receive both the CPAT-R and CPAT interview.

2. Conduct criminal history checks prior to the interview.
    a. If an interview cannot be conducted with a defendant, complete the CPAT-R tool without the interview items (i.e., items 4-11), as well as the clarifying questions for item #5 "Criminal history age," item #8 "Any un-willful FTAs," and item #11 "If yes, please list all charge(s) for prior violent arrest." Note why the defendant could not be interviewed on the front of the form.

3. If the defendant is available for an interview, conduct the CPAT-R interview followed by the original CPAT interview. Any overlapping questions across the CPAT-R and CPAT should be asked only once and recorded for both risk instruments.
    a. Please fill out a hard copy of the CPAT-R.
    b. During the interview, obtain the phone number of a reference who can confirm the information that the defendant has provided. This could be a family member, spouse, friend, or roommate. If the defendant identifies the alleged victim, ask for another reference. Do not contact the alleged victim as a reference

4. Score the CPAT-R and CPAT
    a. CPAT-R:
        i. Call reference to confirm information from items 1 and 2:

            1: Employment/student
            2: Time at current residence
            *Example confirming questions are included on the last page of this document.

        ii. Rely on self-report data for item 3:
            - Do not override based on criminal history records

            3: Problems with alcohol or drugs

        iii. Questions 1-11 have additional boxes to complete if they apply for the defendant. These answers will help inform our validation analysis at the end of the pilot phase of this study.

iv. Use criminal history information from CCIC/NCIC, and Colorado court records to confirm items 4-11:

4: Number of prior arrests
5: Age at first arrest
6: At least 1 arrest in past year
7: Prior FTA
8: FTA within the last year
9: Pending charge at time of arrest
10: Active warrant at time of arrest
11: Prior violence arrest

b. CPAT:

i. Call reference to confirm information from items 1-3:

1: Having a home or cell phone
2: Owning or renting one's residence
3: Contributing to residential payments
*Example confirming questions are included on the last page of this document.

ii. Rely on self-report data for items 4 and 5:

4: Past or current problems with alcohol
5: Past or current mental health treatment

iii. Use criminal history information from CCIC/NCIC, and Colorado court records to confirm items 6-12:

6: Age at first arrest
7: Past jail sentence
8: Past prison sentence
9: Having active warrants
10: Having pending cases
11: Currently on supervision
12: History of revoked bond or supervision

5. Use the **CPAT** score to inform pretrial release recommendation per department protocol.

6. Once this data is recorded, please retain a copy of the risk assessment form(s) in a file for UNC's research team.

CPAT- R Definitions

Employment/education

If the defendant is currently employed, either full or part-time, a student, on disability, a stay-at-home parent or guardian, or retired, they should receive a score of 0-points. This item is intended to address if the defendant is able to work, but currently is not. If he or she is taking classes, either full-time or part-time in vocational school, college, or is unable to work due to disability or other responsibilities (i.e., a primary caretaker) they should receive 0-points. Any type of work, full-time, part-time, or under the table work would count. If unemployed and not represented by one of the above categories, they should receive a score of 2-points. This would include collecting unemployment. Defendants who identify as self-employed should receive a score of 0-points.

If it is confirmed with a reference that the defendant is not: employed, full or part-time, a student, on disability, or retired they should receive a score of 2-points. If there is a discrepancy between the defendant and the reference, the score should be based on the information provided by the reference. If this information cannot be confirmed with a reference, refer to the defendant's unconfirmed response for scoring. If the defendant reported as currently employed or student status, please identify which type the defendant reported in the field on the right of this risk item.

Time at current residence

If the defendant has lived at their current residence for less than one year (365 days) from the day of the CPAT-R interview or does not have a residence, they should receive a score of 1-point. This measure is intended to capture residential stability. They can rent or own, and do not have to be contributing to the payments of the property. This residence is not restricted to Colorado. If the defendant has stable housing in the past year (e.g., homeowner, lease, living with family), but has spent some of this time incarcerated, they should not be given points. Non-permanent mobile homes, hotels, or treatment centers would not count as stable residences.

If it is confirmed with a reference that the defendant has resided at the residence the defendant self-reported for less than one year or does not have a residence they should receive a score of 1-point. If this information cannot be confirmed with a reference, refer to the defendant's unconfirmed response for scoring. If there is a discrepancy between the defendant and the reference, the score should be based on the information provided by the reference. For analysis purposes, please indicate the amount of time the defendant reports living at their residence on the box to the right of the question in months and years. If the defendant has been at there residence for less than one month, please indicate the approximate number of days. If the defendant has no residence or is homeless, indicate this in the field on the right of the question.

Problems with alcohol or drugs

This item captures a self-reported problem with drugs, alcohol or both, either now or in the past. This item should only be scored with 1-point if the defendant self-reported a problem with drugs, alcohol, or both. If the self-reported information pertaining to problems with alcohol or drug use is inconsistent with prior assessments, the defendant's criminal history or current offense, rely on the self-reported response by the defendant for scoring. For analysis purposes, please indicate which condition applies to the defendant by checking the box on the right of this question. In order to gather more information to improve this measure, please indicate if they feel they have had problems with substances, either currently, in the past, or both.

Number of prior arrests

If the defendant had two or more arrests prior to the CPAT interview, they should receive a score of 3-points. If this is their first arrest, or they have only been arrested once prior, they should receive a score of 0-points. Arrests should be counted by date, not by number of prior charge(s). For example, if the defendant was arrested on one day for three charges this counts as one prior arrest. Arrest is defined as booked and/or fingerprinted. If the defendant states that he or she was detained by an officer and immediately released with a summons without being fingerprinted, this is not counted as an arrest. Juvenile arrests are included. Unclassified custodial traffic arrests do not count as arrests.

If the defendant provides conflicting information to their criminal history regarding prior arrests, use official criminal history records for scoring. If there are arrests in different jurisdictions, ensure that the cases are not the same. Indicate the number of prior arrests the defendant has self-reported in the field on the right.

Age at first arrest

If the defendant's age at their first arrest was 20 or younger they should receive a score of 1-point. If they are 21 or older at their first arrest, they should receive a score of 0-points. Arrest is defined as booked and/or fingerprinted. If the defendant states that he or she was detained by an officer and immediately released with a summons without being fingerprinted, this is not counted as an arrest.

If the defendant provides conflicting information to their criminal history regarding age at first arrest, use official criminal history records for scoring. Unclassified custodial traffic arrests do not count as arrests. Please note to the right of the question which age the defendant self-reported and which age the criminal history records indicate.

At least 1 arrest in past year

If the defendant had at least one arrest within one year (365 days) of the CPAT interview, they should receive as score of 3-point. If this is their first arrest or any previous arrests occurred prior to one year from the CPAT interview, they should receive a score of 0-points. Arrest is defined as booked and/or fingerprinted. If the defendant states that he or she was detained by an officer and immediately released with a summons without being fingerprinted, this is not counted as an arrest.

If the defendant provides conflicting information to their criminal history regarding prior arrests, use official criminal history records for scoring. Unclassified custodial traffic arrests do not count as arrests. Indicate if the defendant self-reported an arrest in the past year on the right of the question.

Prior FTAs

Based on Colorado court or CCIC/NCIC records, if the defendant had at least one reported FTA on their record they should receive a score of 3-points. If court records are unavailable, CCIC/NCIC records of prior FTA(s) should be utilized for this risk item.

If the defendant provides conflicting information to their criminal history regarding prior FTA(s), use official criminal history records for scoring. The scorer should include any FTA, even if it is subsequently quashed. Juvenile FTAs count. Civil FTAs do not count. Indicate if the defendant self-reported having at least one prior FTA in the field on the right of the question.

FTA within the last year

Based on Colorado court or CCIC/NCIC records, if the defendant has a reported FTA on their record within one year (365 days) prior to the CPAT interview, they should receive a score of 3-points. If court records are unavailable, CCIC/NCIC records of prior FTA(s) should be utilized for this risk item.

If the defendant provides conflicting information to their criminal history regarding prior FTA(s), use official criminal history records for scoring. The scorer should include any FTA, even if it is subsequently quashed. Juvenile FTAs count. Civil FTAs do not count. Indicate whether the defendant self-reported an FTA within the last year in the field on the right.

> **ALL FTAs, regardless of "willful" or "un-willful" count in this measure. When identifying FTAs within the last year for this survey, please indicate if any FTAs are "un-willful" in the field on the right of this question. Un-willful FTAs are those that records show to have occurred while the defendant is in custody in another jurisdiction.**

Pending charge at time of arrest

At the time of arrest, if the defendant has a pending charge of at least one felony offense, they should be given a score of 2-points. If they have a pending charge, but the charges were misdemeanors only, they should receive a score of 1-point. Traffic and parking violations do not count as a pending charge. Pending charges for municipal and petty offenses do count and should be scored similar to a pending misdemeanor charge with 1-point. If there were no pending charges at the time of their arrest, they should receive a score of 0-points. For example, if a pretrial defendant has three pending charges, one felony and two misdemeanor they would receive 2-points. In contrast, if a pretrial defendant has three pending misdemeanor charges and no pending felony charges they would receive 1-point.

If the defendant provides conflicting information to their criminal history regarding pending charges, use official criminal history records for scoring. A pending charge is defined as an undisposed case for a petty, municipal, misdemeanor, or felony offense that has a scheduled next setting. The exception to this criteria is pending probation revocation. An arrest on a probation revocation warrant that has a scheduled next setting does not count in this category.

Active warrant at time of arrest

If the defendant has a current warrant at the time of their arrest, they should receive a score of 2-points. If there was no active warrant at the time of their arrest, they should be scored a 0-points. Active warrants are defined as a warrant at the time of arrest that does not have a future court date. Active warrants for traffic and parking violations do not count as an active warrant. Active warrants for petty and municipal offenses do count. For example, if a defendant is arrested for an active warrant and additional charge(s) they would score 2-points. In contrast, if a defendant is arrested for an active warrant but no other charge, they would score 0-points.

If the defendant provides conflicting information to their criminal history regarding active warrants, use official criminal history records for scoring. Do not include pending cases for which the defendant is currently on summons or bond for the charges. Probation revocations, Failure to Comply (FTC) Warrants, current FTA warrants, and petty/municipal offense warrants do count in this category. Please indicate if the defendant self-reported having an active warrant in the field on the right of the question.

Prior violent offense arrest

If the defendant has a prior arrest that is considered violent, they should receive a score of 1-point. A violent crime is an offense that a victim is harmed by or threatened with violence.[29] Listed are examples of offenses considered to be violent. This list is not exhaustive so charges not listed or those out-of-state that meet the definition above should also be counted as violent. In addition to the examples provided below, violent arrests may include:

- Used, or possessed and threatened the use of, a deadly weapon; or
- Caused serious bodily injury or death to any other person except another participant;

- Charge examples:
  - Aggravated robbery;
  - Simple robbery;
  - Murder (1 or 2);
  - Manslaughter
  - Criminally negligent homicide
  - Assault (1, 2 or 3);
  - Simple Assault (i.e., municipal violation)
  - Vehicular Homicide
  - Vehicular Assault
  - Kidnapping;
  - First degree arson;
  - First degree burglary;
  - Sexual Assault
  - Sexual Assault on a Child
  - Unlawful Sexual Contact
  - Incest and Aggravated Incest
  - Felony menacing
  - Felony Child Abuse
  - Child Abuse resulting in injury
  - Aggravated cruelty to animals

If the defendant provides conflicting information to their criminal history regarding prior violent offense arrest(s), use official criminal history records for scoring

[29] National Institute of Justice. 2017. "Violent Crimes."

Recommended Question Wording

Interviewing officers are encouraged to use the recommended question wording but may divert from this or use follow-up questions if necessary to obtain reliable assessment information.

1) Employment/education: *Are you currently employed or a student?*

2) Time at current residence: *How long have you lived at your current residence?*

3) Problems with alcohol or drugs: *Do you believe you currently or have ever had a problem with alcohol or drugs?*

4) Number of prior arrests: *Have you ever been arrested before?*

If response is yes: *How many times have you been arrested before?*

If response is no: skip to question 7

5) Age at first arrest: *How old were you when you were first arrested?*

6) At least 1 arrest in the prior year: *How long ago was your most recent arrest before this current arrest?*

7) Prior FTA: *Have you ever failed to return to a scheduled criminal court date?*

Clarifying Question: *Have you ever had an FTA?*

If response is no: skip to question 9

8) FTA within the prior year: *When was the last time you failed to return to court?*

9) Pending charge at the time of arrest: *Besides what you were currently arrested for, do you have any other pending criminal court cases?*

10) Active warrant at arrest: *Did you have any active warrants at the time of your arrest?*

11) Prior violent offense: *Have you ever been arrested for a violent crime?*

Recommended Information Confirmation Questions

1) **Own a home or cell phone (From Original CPAT):**

Does [the defendant] have a working phone number for you to contact him/her?

2) **Employment/education (From CPAT-R):**

Do you know where [the defendant] works?

Are they currently a student?

Are you aware if they on receiving any disability assistance?

2) **Residence Questions from CPAT and CPAT-R:**

Do you know what city [the defendant] currently lives in?

How long have they lived there?

Have they moved in the previous year?

Do they rent or own their current residence?

Do they contribute to the [rent or mortgage]?

APPENDIX G

CPAT-R Pilot Feedback Survey

1. During the CPAT-R pilot study (Nov 13, 2019 to Feb 13, 2019) did you conduct CPAT-R interviews? Y/N

1a. (No) Describe your familiarity with the CPAT-R. (skip to 8)

1b. (Yes) On an average day, what percentage of the pretrial risk assessment interviews that you conducted were CPAT-R interviews? 100 – 0%

3. How long would you say an average CPAT-R interview would take in minutes: _____

4. Are there any risk items on the CPAT-R that you found difficult to solicit information about during interviews with pretrial defendants? *link to CPAT-R risk items.

5. Did you use the recommended question wording from the CPAT-R Pilot Implementation Guide? Yes/No

5a. (Yes) skip to 6.

5b. (No): Describe how you modified the recommended question wording to solicit information during interviews.

6. How long would you say an average CPAT-R investigation would take in minutes: _____

7. During the CPAT-R investigation, did you attempt to confirm information with a reference?

7a. (No) skip to 8

7b. (Yes) On an average day, what percentage of interviews were you successful at contacting a pretrial defendant's reference? 100 – 0%

7c. Did you use the recommended question wording from the CPAT-R Pilot Implementation Guide to confirm information? Yes/No

7d. (Yes) skip to 7f

7e. (No): Describe how you modified the recommended question wording to solicit information when confirming information.

7f. On an average day, what percentage of interviews were you able to successfully confirm a pretrial defendant's information? 100 – 0%

8. Are there items on the CPAT-R that you feel like should be eliminated? If so, which items and why? *link to CPAT-R items

9. Are there items on the CPAT-R that you feel should be added? If so, why these items? *link to CPAT-R items

10. Are there any definitions of terms used for the CPAT-R that you feel are unclear? If so, please describe. *link to CPAT-R item definitions

11. What are your thoughts on the investigation portion of the CPAT-R overall?

12. What are your thoughts on confirming interview information, overall?

13. What resources or training do you think would be beneficial for working with the CPAT-R in addition to the pilot implementation guide?

14. What do you like about the CPAT-R?

15. What do you feel could be improved with the CPAT-R?

16. What county do you currently work in?

17. How long have you worked in your current position?

18. What is your current position?

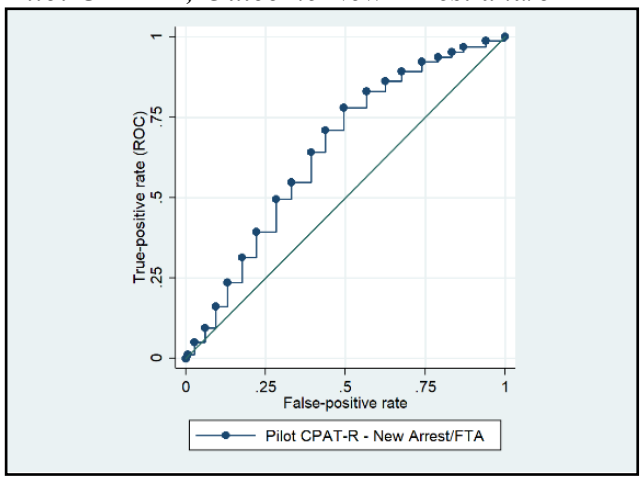
19. Do you have any additional feedback?

APPENDIX H

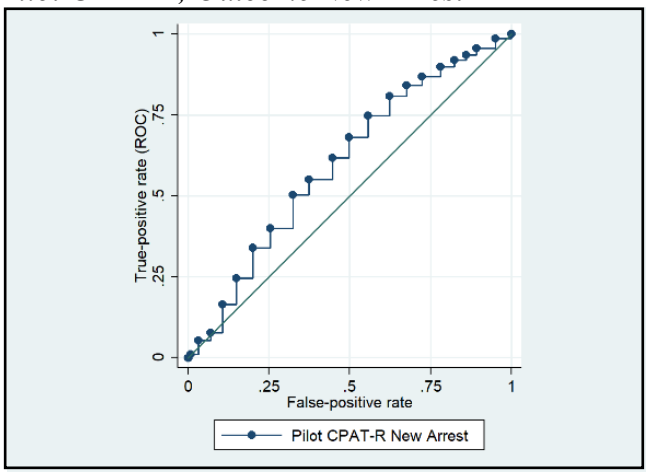
ROC Curves and Calibration Plots – Phase 3 Validation

ROC Curves

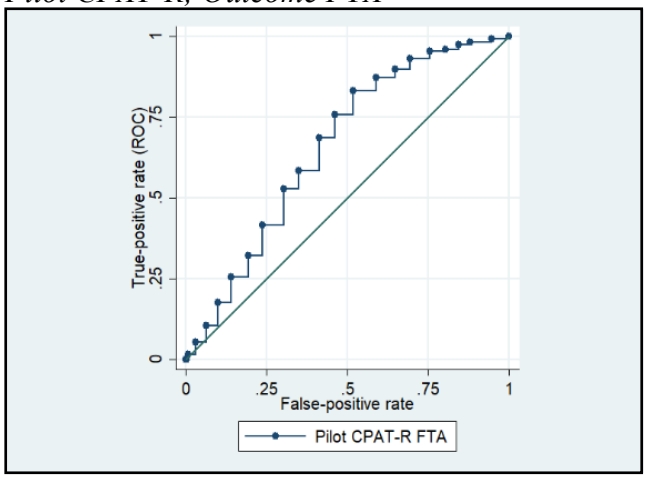
Pilot CPAT-R; Outcome New Arrest and/or FTA



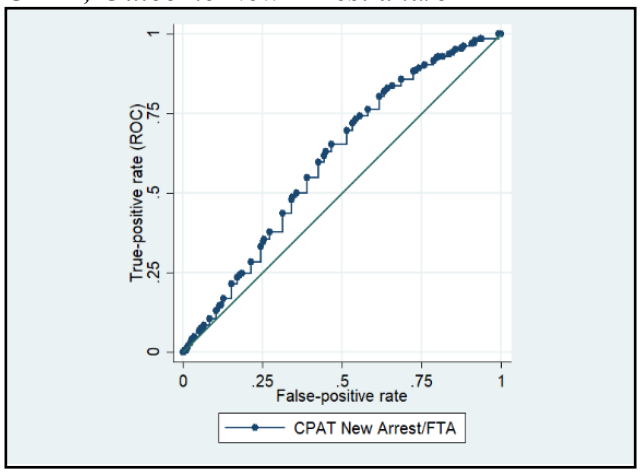
Pilot CPAT-R; Outcome New Arrest



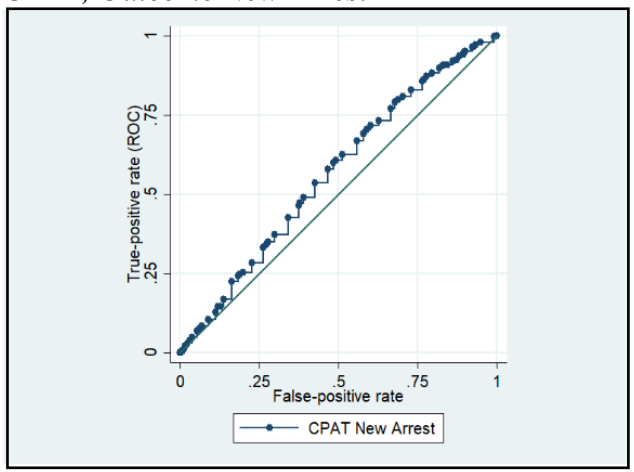
Pilot CPAT-R; Outcome FTA



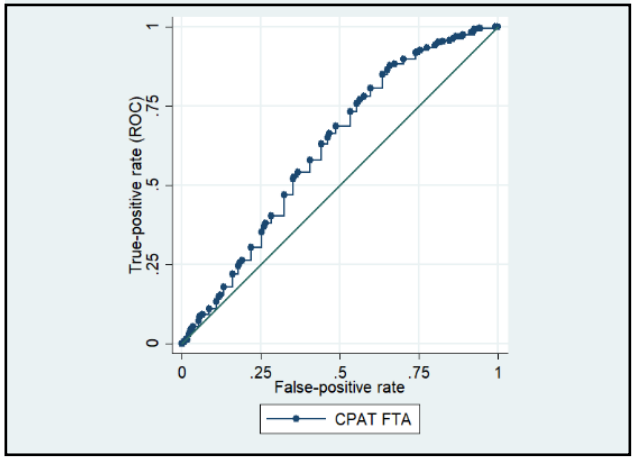
CPAT; Outcome New Arrest and/or FTA



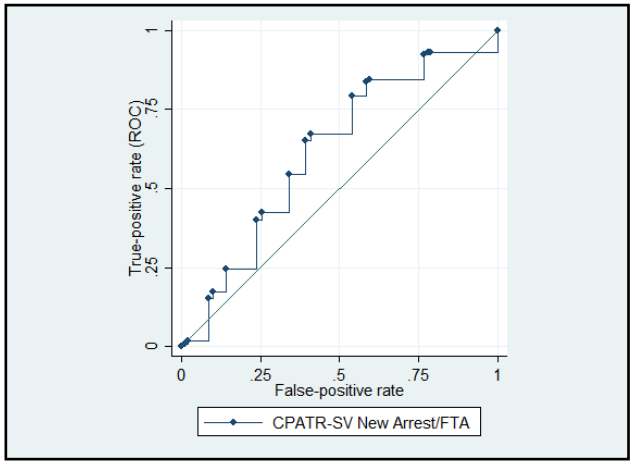
CPAT; Outcome New Arrest



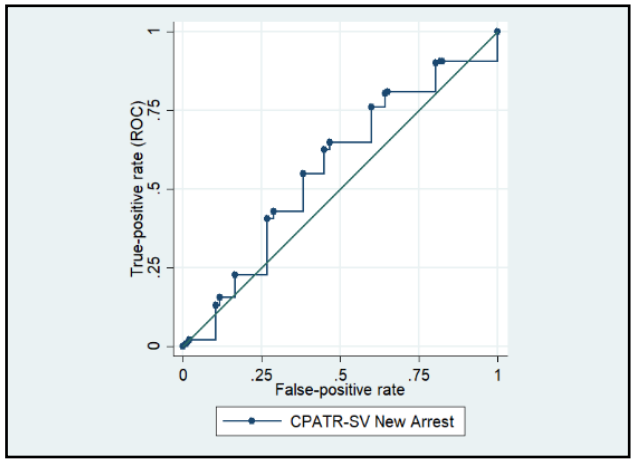
CPAT; Outcome FTA



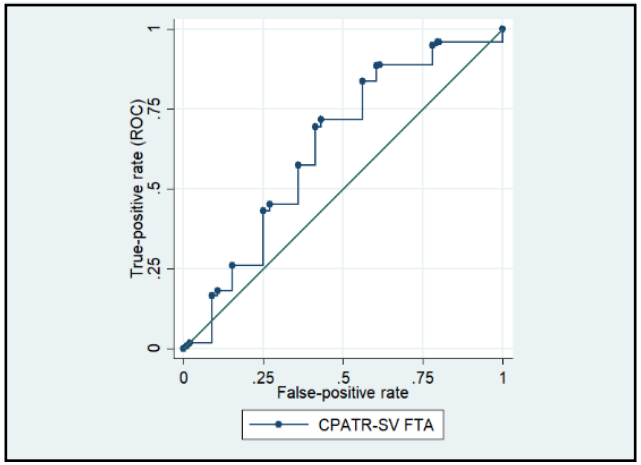
CPATR-SV; Outcome New Arrest/FTA



CPATR-SV Outcome New Arrest

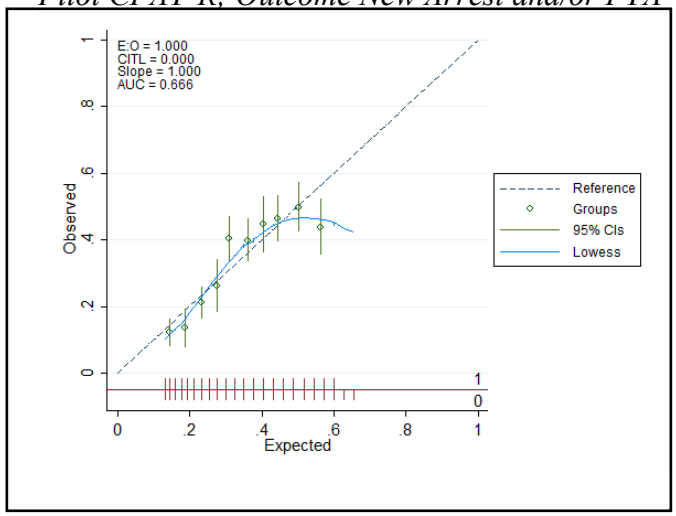


CPATR-SV; Outcome FTA

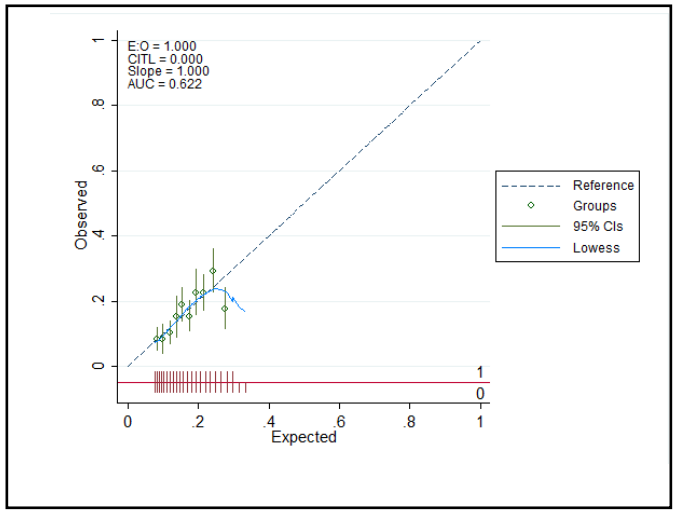


Calibration Plots

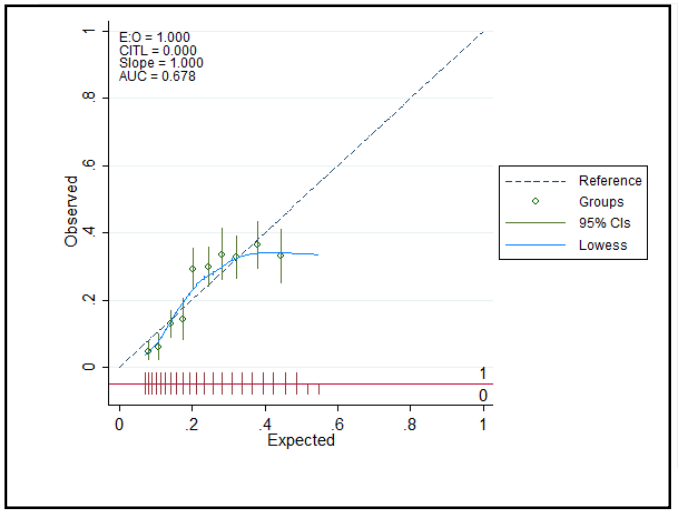
Pilot CPAT-R; Outcome New Arrest and/or FTA



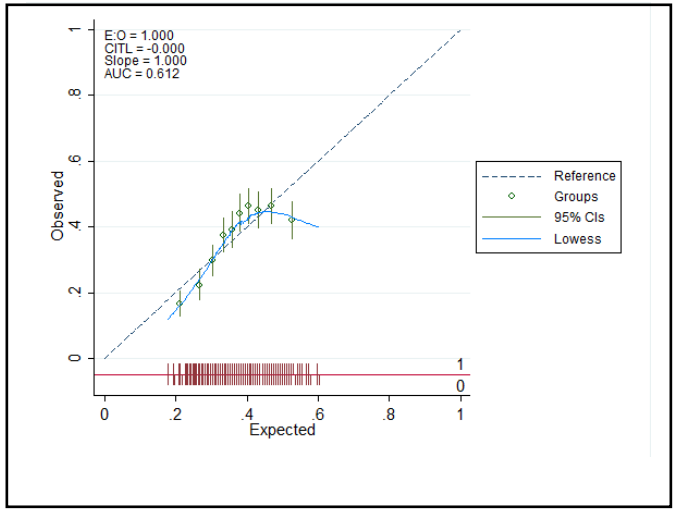
Pilot CPAT-R; Outcome New Arrest



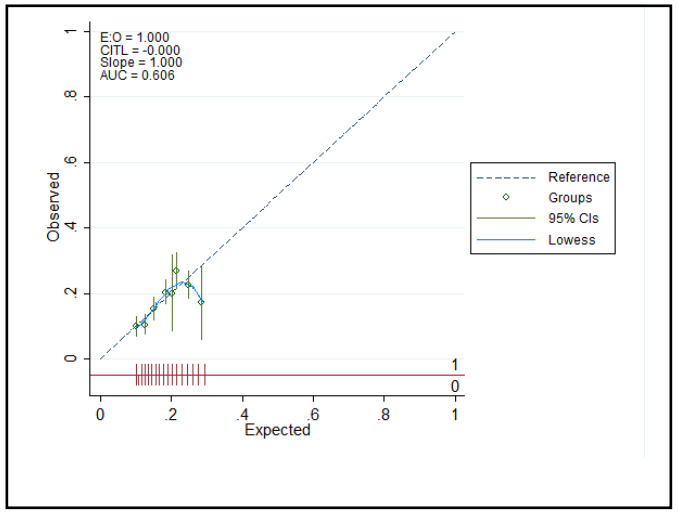
Pilot CPAT-R; Outcome FTA



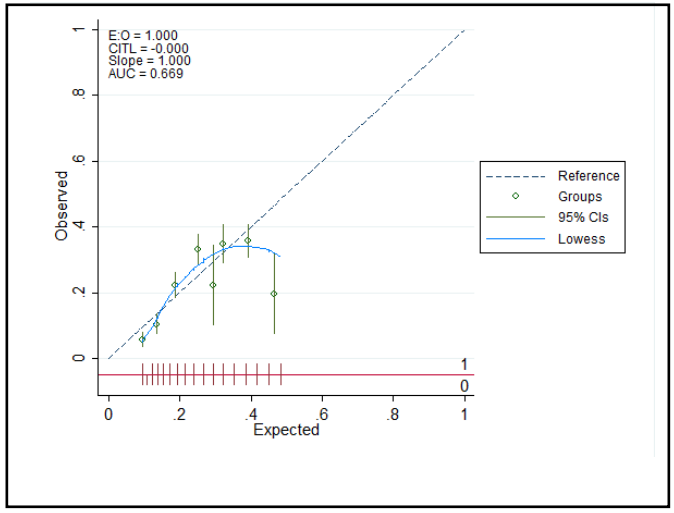
CPAT; Outcome New Arrest/FTA



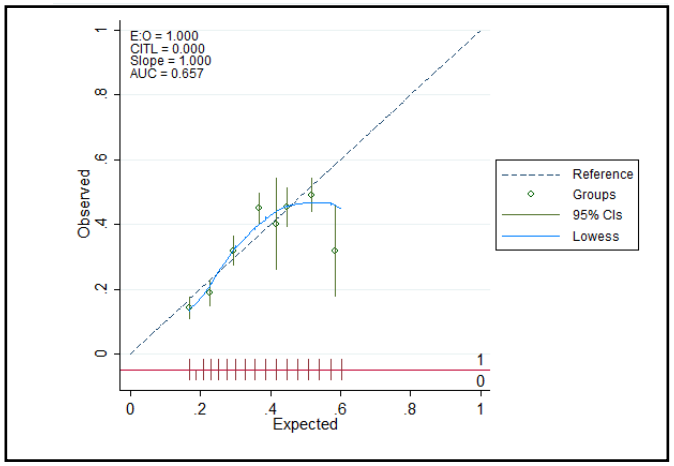
CPAT; Outcome New Arrest



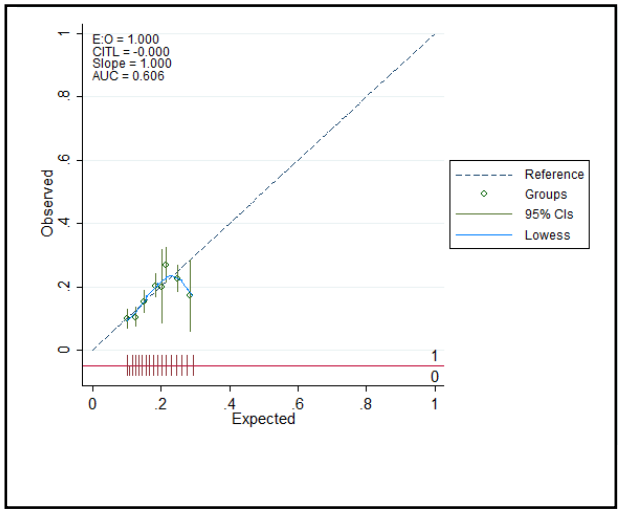
CPAT; Outcome FTA



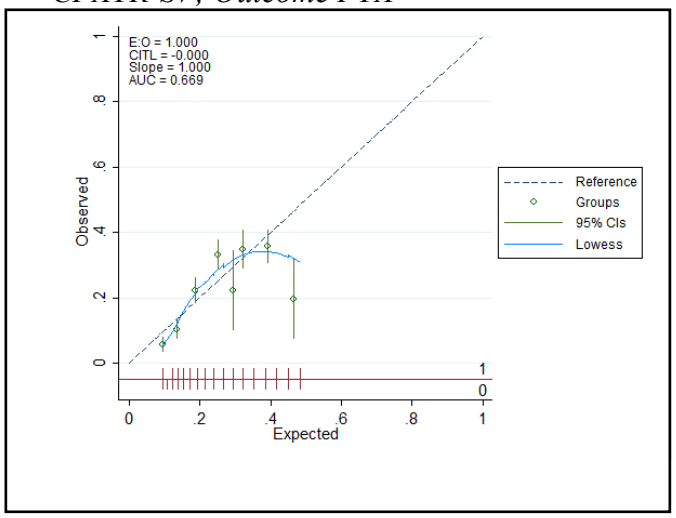
CPATR-SV; Outcome New Arrest/FTA



CPATR-SV; Outcome New Arrest



CPATR-SV; Outcome FTA



APPENDIX I

Risk Item Level Assessment Errors Across Sub-groups

Race/ethnicity: Validation sample

	False Positive Rate			**False Negative Rate**			**Failure Prediction Error**			**Success Prediction Error**			**Overall Prediction Error**		
	W	B	H	W	B	H	W	B	H	W	B	H	W	B	H
Pilot CPAT-R	.24	.37	.29	.52	.55	.48	.52	.66	.50	.23	.27	.27	.41	.50	.43
Risk Factor															
Employment/education	.34	.47	.42	.55	.57	.44	.63	.72.	58	.27	.32	.29	.41	.50	.43
Time at current residence	.34	.32	.26	.61	.66	.71	.66	.68	.62	.29	.29	.34	.42	.42	.42
Problems with alcohol or drugs	.42	.29	.34	.44	.63	.64	.63	.38	.63	.25	.29	.35	.43	.42	.45
Prior arrest	.46	.37	.46	.44	.55	.52	.65	.66	.64	.27	.27	.34	.46	.42	.48
Arrest in the last year	.40	.50	.48	.33	.32	.38	.58	.63	.58	.20	.22	.28	.38	.45	.44
Age at first arrest	.52	.67	.67	.32	.30	.26	.63	.69	.62	.23	.28	.30	.46	.56	.52
Prior FTA	.55	.78	.67	.23	.11	.16	.62	.67	.60	.19	.17	.21	.45	.58	.49
FTA in the last year	.23	.37	.31	.51	.52	.44	.51	.64	.51	.23	.26	.26	.32	.41	.26
Pending charge at arrest	.13	.04	.13	.84	.91	.90	.64	.50	.70	.30	.29	.36	.35	.30	.40
Active warrant	.22	.38	.34	.60	.54	.52	.55	.65	.56	.26	.27	.30	.34	.42	.40
Prior violent arrest	.35	.68	.53	.45	.18	.39	.59	.66	.62	.24	.20	.32	.38	.53	.38

a) White, *n* = 656, new arrest/FTA baserate = .31; Black, *n* = 186, new arrest/FTA baserate = .30;
Hispanic, *n* = 331, new arrest/FTA baserate =.35

Residential status: Validation Sample

	False Positive Rate		**False Negative Rate**		**Failure Prediction Error**		**Success Prediction Error**		**Overall Prediction Error**	
	H	NH	H	NH	H	NH	H	NH	H	NH
No prior violent modification	.60	.31	.26	.45	.51	.57	.33	.22	.45	.35
Risk Factor										
Employment/Education	.74	.34	.29	.57	.54	.65	.46	.27	.54	.41
Time at Current Residence	.97	.22	.00	.79	.56	.71	.00	.30	.55	.39
Problems with alcohol or drugs	.48	.38	.42	.55	.52	.66	.38	.27	.45	.43
Prior arrest	.48	.46	.40	.47	.51	.67	.37	.27	.45	.47
Arrest in the last year	.57	.42	.28	.37	.51	.61	.33	.22	.45	.41
Age at first arrest	.60	.60	.35	.29	.55	.66	.40	.24	.49	.50
Prior FTA	.72	.60	.15	.20	.53	.64	.28	.18	.47	.48
FTA in the last year	.38	.26	.40	.51	.45	.55	.33	.23	.39	.34
Pending charge at arrest	.10	.13	.88	.85	.52	.67	.43	.50	.44	.35
Active Warrant	.44	.26	.51	.59	.54	.59	.41	.25	.47	.36

H = Homeless, NH = Housed

a) Homeless, *n* = 206, new arrest/FTA baserate = .43; Housed, *n* = 1,060, new arrest/FTA baserate = .30

APPENDIX J

CPAT & CPATR-SV: Accuracy Equity and Predictive Parity Assessment

Full Sample - CPAT Assessment Errors

	False Positive Rate	False Negative Rate	Failure Prediction Error	Success Prediction Error	Overall Prediction Error
Race/Eehnicity[a]					
White	.53	.31	.58	.27	.45
Black	.64	.23	.63	.24	.51
Hispanic	.52	.28	.58	.33	.47
Sex[b]					
Male	.54	.30	.45	.28	.45
Female	.47	.36	.42	.32	.43
Residential Status[c]					
Homeless	.79	.16	.51	.42	.49
Housed	.51	.35	.62	.27	.46

a) White, *n* = 2,317, new arrest/FTA baserate = .35; Black = 499, new arrest/FTA baserate = .33; Hispanic, *n* = 733, new arrest/FTA baserate = .38
b) Male, *n* = 2,636, new arrest/FTA baserate = .37, Female = 807, new arrest/FTA baserate = .40
c) Homeless, *n* = 571, new arrest/FTA baserate = .48; Housed = 3,186, new arrest/FTA baserate = .33

Full Sample - CPATR-SV Assessment Errors

	False Positive Rate	False Negative Rate	Failure Prediction Error	Success Prediction Error	Overall Prediction Error
Race/ethnicity[a]					
White	.75	.09	.60	.16	.51
Black	.73	.14	.63	.20	.53
Hispanic	.66	.15	.56	.22	.47
Sex[b]					
Male	.73	.11	.58	.19	.50
Female	.72	.13	.55	.24	.48
Residential Status[c]					
Homeless	.80	.09	.49	.29	.46
Housed	.72	.12	.62	.18	.52

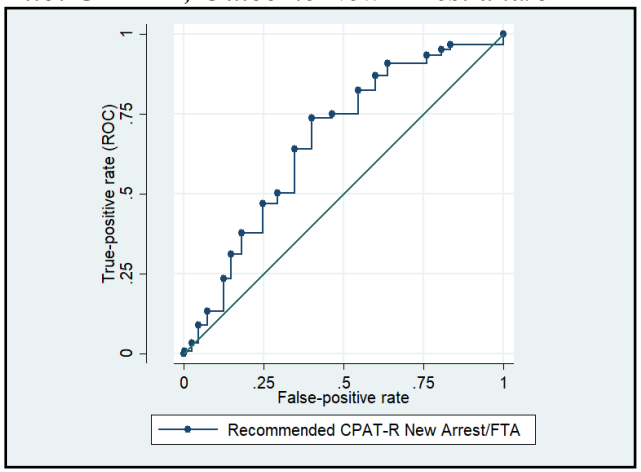
a) White, *n* = 2,317, new arrest/FTA baserate = .35; Black, *n* = 499, new arrest/FTA baserate = .33; Hispanic, *n* = 733, new arrest/FTA baserate = .38
b) Male, *n* = 2,636, new arrest/FTA baserate = .37; Female, *n* = 807, new arrest/FTA baserate = .40c) Homeless, *n* = 571, new arrest/FTA baserate = .48; Housed, *n* = 3,186, new arrest/FTA baserate = .33

APPENDIX K

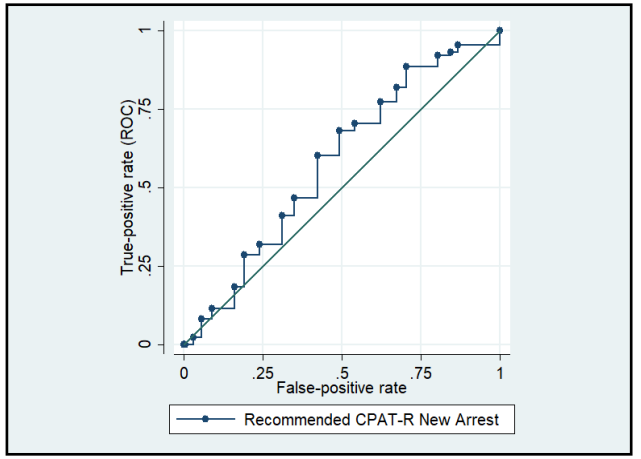
ROC Curves and Calibration Plots – Recommended Tool

ROC Curves

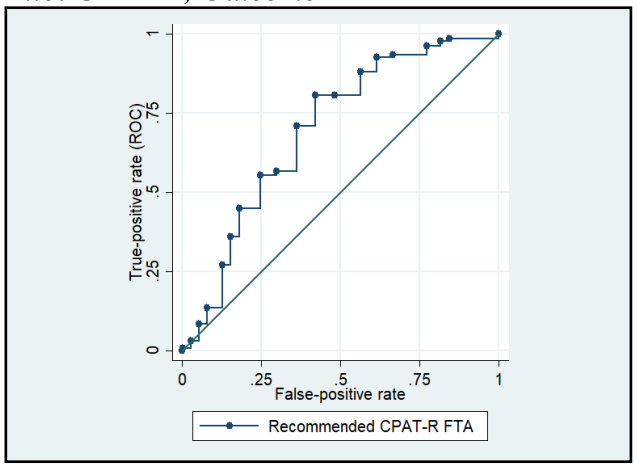
Pilot CPAT-R; Outcome New Arrest and/or FTA



Pilot CPAT-R; Outcome New Arrest

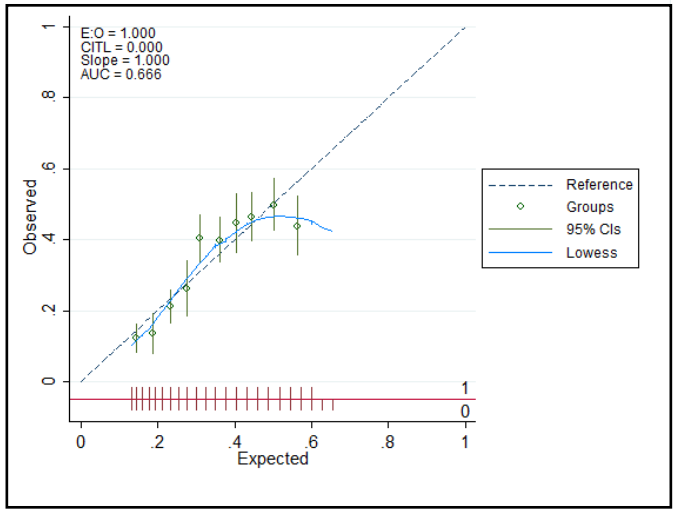


Pilot CPAT-R; Outcome FTA

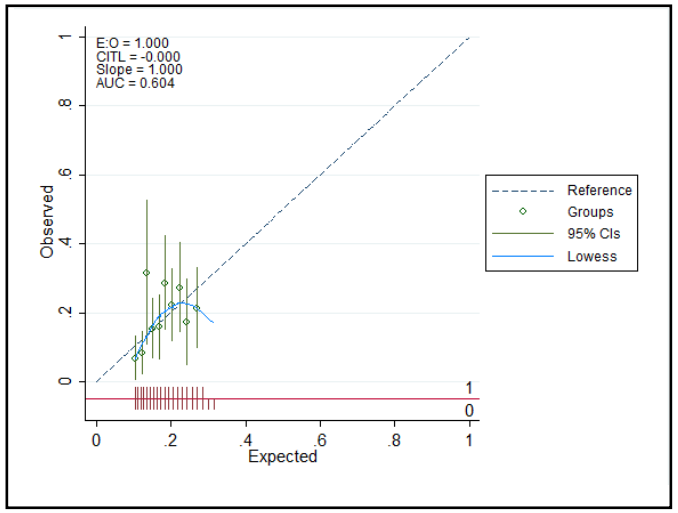


Calibration Plots

Pilot CPAT-R; Outcome New Arrest and/or FTA



Pilot CPAT-R; Outcome New Arrest



Pilot CPAT-R; Outcome FTA