

RESEARCH REPORT

Performance-Based Contracting for Colorado Community Corrections

Rayanne Hawkins
October 2020

Ashlin Oglesby-Neal

Libby Doyle

Isabella Remor



ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.

Contents

Acknowledgments	iv
Executive Summary	v
Performance-Based Contracting for Colorado Community Corrections	1
Background	2
Summary of Colorado's 2015 PBC Plan	3
Performance-Based Contracting Payment Models	7
Recommendations for Risk-Adjusted Outcome Measures	11
Recommendations for PBC Payment Models in Colorado	15
Recommendations for the PBC Implementation Plan	23
Conclusion	25
Appendix A. Technical Appendix	27
Notes	47
References	48
About the Authors	49
Statement of Independence	50

Acknowledgments

This report was funded by the Colorado Department of Public Safety, Division of Criminal Justice, Office of Community Corrections. We appreciate their collaboration throughout this project. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

We are thankful for the support and participation of many Colorado community corrections stakeholders, including Governor's Community Corrections Advisory Council members, community corrections board members, and program administrators and staff.

Executive Summary

The Urban Institute was contracted by the Colorado Office of Community Corrections (OCC) to help it develop a set of risk-adjusted outcome measures, a proposed performance-based contracting (PBC) payment model, and an implementation plan for transitioning community corrections program contracts from fee-for-service to PBC contracts. To meet these goals, Urban conducted document reviews, a scan of PBC practices, and a survey of and interviews with community corrections stakeholders from OCC, judicial districts, probation and parole, and program managers and frontline staff. Urban generated a series of recommendations that we have broken into three categories: risk-adjusted outcome measures, PBC models, and implementation.

Risk-Adjusted Outcome Measures

Urban analyzed data from OCC and the Colorado state court system, surveys, and interviews to develop six outcome measures that can be used to determine the success of community corrections clients. There are two positive outcomes (successful completion and employment gain or retention) and three negative outcomes (escapes, technical violations, and recidivism). Another outcome is Level of Supervision Inventory (LSI) score change.

Urban then developed a method for adjusting the outcomes for risk. Urban developed two categories of risk: low/medium and high/very high. Those categories were used to calculate the five-year state average baseline outcomes for each risk level. Urban then used program-level data on clients to determine whether each program was high/very high risk, defined as having more than 50 percent of clients in a year with LSI scores of high or very high. The remaining programs (i.e., those for which 50 percent of more of clients had LSI scores of low or medium risk) were determined to be low/medium risk.

PBC Models

The PBC model should include metrics for public safety and security (measured by Core Security Audits [Core]), use of evidence-based practices (measured by Program Assessment for Correctional Excellence [PACE]), and the risk-adjusted outcome measures developed by Urban. This report includes two potential PBC models, one gradual (table 1) and another with lower baseline payments and higher

potential incentive rewards (table 2). Both rely on a per diem payment model where the base per diem is earned regardless of performance. Because compliance with PACE and Core are requirements of all programs, the only bonus to be earned in each model is for superior performance on the risk-adjusted outcome measures.

TABLE 1
Gradual Transition Model

	FY 22 (7/1/21– 6/30/22)	FY 23 (7/1/22– 6/30/23)	FY 24 (7/1/23– 6/30/24)	FY 25 (7/1/24– 6/30/25)	FY 26 (7/1/25– 6/30/26)
Base per diem	100%	99%	98%	97%	97%
Completion	baselines	1%*	1%*	1%*	1%*
Recidivism	baselines	1%*	1%*	1%*	1%*
PACE or Core	baselines	baselines	2%	3%	3%
Max payment	100%	101%	102%	102%	102%

Note: * payment based on data from previous one to two fiscal years.

TABLE 2
Progressive Risk and Reward Model

	FY 22 (7/1/21– 6/30/22)	FY 23 (7/1/22– 6/30/23)	FY 24 (7/1/23– 6/30/24)	FY 25 (7/1/24– 6/30/25)	FY 26 (7/1/25– 6/30/26)
Base per diem	100%	98%	96%	94%	92%
Completion	baselines	2%*	2%*	3%*	4%*
Recidivism	baselines	2%*	2%*	3%*	4%*
PACE or Core	baselines	baselines	4%	6%	8%
Max Payment	100%	102%	104%	106%	108%

Note: * payment based on data from previous one to two fiscal years.

To better understand the models, we present the following example of the gradual transition model in practice. Assuming that the FY 2024 per diem is \$50.00 per day for a regular community corrections bed, every program would receive 98 percent of \$50.00, or \$49.00, for every client served regardless of the program’s performance on the metrics. If the program met its performance targets on either the PACE or Core (assuming it received one of the two assessments each year), it would get another \$1.00 per client per day. If the program met or exceeded the target for successful completion, it would receive an additional \$0.50 per client per day. Similarly, if it performed well on the recidivism metric, it would get another \$0.50 per client per day, making \$51.00 the maximum payment for the gradual model in FY 2024 provided the program did well on the measures for all three metrics (i.e., the PACE or Core assessment, successful completion, and recidivism). If the program did well on the PACE or Core and one of the outcome metrics, it would get \$50.50 per client per day.

The models have the same benefits, which are as follows:

- **Gradual adjustments.** The tables include a gradual reduction in the per diem, and in the first year, programs will get full payment. This should allow programs to become familiar with the performance metrics and focus on taking needed steps to improve their scores without risk of reduced funding.
- **Annual adjustments.** The state budget committee and OCC can increase or decrease the per diem rate based on the current state budget and fiscal environment.
- **Flexible targets.** The rate table offers a flexible structure where OCC can assess programs' performance on risk-adjusted outcomes as compared with the state average and programs' year-over-year improvement on PACE and Core.
- **Flexible schedule.** Allowing programs to earn the incentive by improving on either the PACE or Core assessment affords the state flexibility for changing how often each assessment is administered.
- **Minimal contract changes.** This new payment structure can be incorporated in judicial district and program contracts by simply amending the letter that includes the fiscal year's bed allocation and per diem amount because it still relies on a per diem payment structure.

Moreover, OCC will need to make the five following decisions related to the PBC model to finalize the PBC implementation plan:

- **Select a model.** The office needs to select one of the two PBC payment models based on the level of incentive the state can afford and whether programs can accept a lower base per diem rate without negatively impacting the quality of their services.
- **Select two outcomes from the six.** Urban proposed six outcomes that OCC should track. Because it is difficult to focus on six outcomes and because using all six would reduce the percentage assigned to each incentive, Urban suggests selecting the two most important outcomes and using them in the model.
- **Create a warning system.** An important part of a performance measurement system is creating nonfinancial incentives (such as a penalty for failure) to improve metrics multiple years in a row. For example, the Pennsylvania Department of Corrections can cancel community corrections contracts if recidivism scores are poor in two consecutive measurement cycles.
- **Choose baseline targets for PACE and Core.** The Office of Community Corrections needs to decide whether the baseline measures will be based on averages across programs or on an individual program's year-over-year improvement on PACE and Core.

- **Determine the frequency of PACE and Core.** The plan is for OCC to administer either PACE or Core assessments to each program in a two-year cycle so that programs get one or the other annually. Two years may not be enough time to improve scores, but more data are needed to know the potential for score improvement in a two-year period.

Implementation

Urban has four main recommendations regarding PBC implementation. First, OCC should make sure that the future data system collects all the needed information for PBC and that data from programs are entered in an accurate and timely fashion. Second, it should continue to offer training to programs on evidence-based practices and other ways to improve safety and outcomes for clients. Third, it should consider simplifying PACE and Core so that they are much easier to administer and include the most important PACE factors and Core standards. Fourth, OCC should find a way to transition gradually to the PBC payment model.

Conclusion

Urban has provided OCC with six risk-adjusted outcome measures, two PBC models, and four main recommendations for implementation. The office will need to work with the Governor's Community Corrections Advisory Council and the joint budget committee to select the PBC model that makes the most sense for the state's priorities and current fiscal environment.

Performance-Based Contracting for Colorado Community Corrections

The Urban Institute was contracted to advise the Colorado Office of Community Corrections (OCC) in refining its plan to transition community corrections programs from fee-for-service to performance-based contracting (PBC). In this report, we build off work outlined by the Governor's Community Corrections Advisory Council's (2015) *Performance Based Contracting (PBC) for Colorado Community Corrections: General Attributes of a PBC Plan*, hereafter referred to as the 2015 PBC plan. This report includes recommendations on three key elements introduced in the 2015 PBC plan: calculating risk-adjusted outcome measures, creating a PBC payment model, and presenting considerations for implementation of performance-based contracting.

To develop risk-informed outcomes and recommendations for a PBC implementation plan, Urban researchers employed quantitative and qualitative methods. To gain a better understanding of relevant measures of program performance and current program review processes, Urban conducted semistructured in-person and phone interviews and fielded a short survey with community corrections staff and stakeholders. The information gathered through these activities was used to guide the selection of risk-informed outcomes as well as the logistics of a PBC implementation plan. To develop these outcomes and baseline measures, Urban collected individual-level client program data and recidivism information. The administrative program data come from OCC's Community Corrections Information and Billing system and the recidivism data come from the Colorado Judicial Department. Urban researchers also collected program-level assessment and evaluation data from the Program Assessment for Correctional Excellence (PACE) and Core Security Audit (Core) baseline scores. Urban worked closely with OCC to use the program- and client-level data to develop a PBC implementation plan. The Urban research team also reviewed PBC models in other states and research on best practices in PBC implementation.

We begin this report with background on the PBC progress to date in Colorado and a summary of Colorado's 2015 PBC plan. We then highlight relevant PBC payment models. Lastly, we present recommendations on three key elements: the risk-adjustment outcome calculation methods, two potential PBC payment models, and advice around an implementation plan for PBC.

Background

Colorado has 33 residential and nonresidential community corrections programs that provide core community corrections supervision services and handle specialized cases. They function both as a step in the reentry process for people returning to the community from prison and as a diversion program for people sent there in lieu of prison. The Office of Community Corrections (a part of the Colorado Department of Public Safety, Division of Criminal Justice) oversees the funding and evaluation of services for community corrections. Services are currently funded using a per diem-based, fee-for-service payment model. To better align payment for community corrections operations and services with the state's emphasis on evidence-based practices (EBPs) and improvements in outcomes, the state is considering supplementing the per diem with incentives from a performance-based contracting (PBC) payment model.

In August 2015, the Colorado Governor's Community Corrections Advisory Council approved *Performance Based Contracting (PBC) for Colorado Community Corrections: General Attributes of a PBC Plan* (2015 PBC plan). It highlighted three areas of performance: the adoption of EBPs, safety and security, and risk-informed outcomes. The plan called for assessments of all three performance areas and the creation of a payment model that would incentivize performance improvements in each area. The 2015 PBC plan laid the groundwork for the next steps for implementing PBC, which included developing an assessment tool to measure adherence to EBPs, updating the community corrections standards, developing a tool for assessing performance against the safety standards, and hiring a firm to develop risk-informed outcome measures, calculate the baseline measures, and propose a payment model and an implementation plan for PBC.

In 2017, the Colorado Community Corrections Standards, which define operational and security standards for community corrections programs, were updated. That same year, PACE, which assesses the use of EBPs, was administered to the first community corrections program. Then, in 2018, administration processes for Core, which assesses programs' adherence to the Colorado Community Corrections Standards, were updated. Lastly, in 2019, the Urban Institute was contracted to consult on the establishment of risk-adjusted outcomes measures, provide examples of PBC in practice, develop a PBC payment model, and provide recommendations for implementing performance-based contracting. Throughout this report, we refer to Colorado's current and desired contract as a performance-based contract, although in the section on PBC payment models, we acknowledge that it is more narrowly an outcomes-based contract.

Summary of Colorado's 2015 PBC Plan

The 2015 PBC plan described in this section was considered tentative, and OCC planned to revisit it once it had hired a consultant to help it develop risk-adjusted outcome measures and a PBC implementation plan. The 2015 plan outlined seven principles for PBC, which are as follows:

- Principle 1: the PBC model should be forward looking and based on long-term goals rather than the present state of affairs.
- Principle 2: the model should not face barriers to adoption from historical or existing practices.
- Principle 3: the model should be used to reward outstanding performance, include mechanisms for addressing poor performance, and remain nonpunitive.
- Principle 4: the PBC model should include performance metrics for assessing security, risk-informed outcome measures, and use of EBPs.
- Principle 5: the model and its metrics should be easily understood and fair to all community corrections stakeholders of OCC.
- Principle 6: the model should incorporate the partnership between state and local partners.
- Principle 7: the model's baseline results should leave room for improvement and any baseline score reports should acknowledge this.

In addition to the principles, the 2015 plan proposed using a composite score of a provider's PACE score, Core score, and risk-adjusted outcome measures to determine its level of per diem reimbursement. The following sections explain each of these assessments and the corresponding PBC payment system.

PACE

The PACE tool evaluates how well community corrections programs adhere to the Colorado Community Corrections Standards. The standards incorporate the National Institute of Corrections' Principles of Effective Intervention, which aim to improve community corrections outcomes, such as risk reduction, successful program completion, and recidivism. PACE rates each community corrections program on seven factors to determine fidelity to the National Institute of Corrections' EBPs as well as each program's capacity to implement such practices. Providers are rated on a scale of zero (0) to four (4), and are reassessed every two years. PACE uses staff audio recordings, staff observations and interviews, client case file reviews, reviews of client handbooks and program policy and procedure, and

surveys of staff and clients to better understand each specific program's competency and adherence to the EBPs promoted by the National Institute of Corrections and OCC.

The factors that the PACE tool measures are under the direct control of the provider, meaning that each provider can improve its performance by improving its practices. PACE measures practices based on the following factors:

- **Factor 1: Assess Actuarial Risk and Need**
 - » Risk assessments are conducted well and explained to clients, including normative feedback on their score.

- **Factor 2: Enhance Intrinsic Motivation**
 - » Motivational interviewing skills are demonstrated and used with clients in a variety of contexts.

- **Factor 3: Target Intervention/Responsivity**
 - » Individual client differences are taken into account in prioritizing criminogenic needs and how clients are engaged in program work.

- **Factor 4: Skill Train with Directed Practice**
 - » Staff clarify their different roles with clients regularly and facilitate skills practice that will help clients address criminogenic needs.

- **Factor 5: Increase Positive Reinforcement**
 - » A level system, a contingency management system, and specific verbal praise are used in the program to reward behavioral progress and stabilization.

- **Factor 6: Respond to Violation Behavior with Effective Practices**
 - » Terminations, sanctions, and behavioral interventions adhere to principles of procedural justice (i.e., swift, fair, consistent, proportional, parsimonious, and transparent).

- **Factor 7: Engage Ongoing Support in Natural Communities**
 - » Staff help clients identify and engage prosocial support systems in the community, encourage participation in prosocial community involvement, and support exposure to prosocial support networks by hosting organized group activities at the program.

To complete PACE, OCC staff request documentation, review the files, and conduct an on-site visit to collect data, which are then used to create a PACE profile for the program. The results are then presented during interactive profile review meetings, typically two to six weeks after the on-site evaluations. The entire process takes four to five months. Additionally, OCC also offers training and technical assistance on an ongoing basis.

In addition, OCC sought to develop PACE baseline scores for all community corrections programs before implementing PBC in community corrections. In April 2018, PACE relaunched after significant OCC staff turnover, and OCC began evaluating programs, averaging one program evaluation every six weeks. COVID-19 interfered with the schedule for administering PACE, and at the time of publication, 29 of 33 baseline PACE evaluations had been conducted and the first baseline assessment had a mean score of 1.55 out of 4.

Core Security Audits

In addition to the PACE evaluation tool, OCC performs Core Security Audits on every program to assess how well each complies with community corrections security standards. There are 94 community corrections standards, but Core focuses on 21, 14 of which are scored and all of which include a public safety component related to client supervision. The audit process itself looks at a large number of program data sources, including observations of staff and client interactions, staff and client interviews, camera reviews, and reviews of client case files and program logs. Providers are rated on a scale of zero (0) to three (3), and are reassessed every two years. Similar to PACE, an audit report is produced based on the collected data and is then brought back to an interactive in-person meeting with program stakeholders. Like the PACE factors, the Core standards that are measured are under the direct control of the providers, allowing them to improve their practices based on their evaluations.

Baseline Core scores are in progress. The original goal was to complete security audits for all programs by June 30, 2020. Office of Community Corrections staff began performing audits in February 2019 and had an audit scheduled every other week to meet the deadline. However, because the COVID-19 pandemic has restricted in-person site visits, this schedule has shifted. At the time of this writing, 24 of 30 Core audits had been completed.

Risk-Informed Outcome Measures

While completing PACE and Core, the State of Colorado has hired Urban to calculate risk-informed or risk-adjusted outcome measures. The 2015 PBC plan suggested that OCC consider measures for four risk-informed measurements: recidivism, success rate, new crime rate, and escape rate, all adjusted by client risk level. The plan proposed adjusting each measure by the average Level of Supervision Inventory (LSI) risk score of clients in the program. Risk adjustment was identified as crucial because client outcomes vary by risk level and programs that have higher-risk populations should not be penalized.

Composite Measure

The 2015 PBC plan stipulated that scores from PACE, Core, and the risk-adjusted outcomes be weighted and combined to determine a composite score from which per diem levels would be determined. The Office of Community Corrections believes that these three performance measures represent its general definition of public safety and are demonstrative of overall program performance. Programs' overall performance ratings were intended to have the following weighting: 50 percent from PACE, 25 to 40 percent from Core, and 10 to 25 percent from risk-adjusted outcomes.

Program Performance and Payment

Colorado's 2015 plan featured six (6) performance-banded program categories that would be determined by a program's overall performance score. These categories reflected the level of funding that the programs would receive from the state. Each "type" of program would be incentivized to achieve the next highest type. Types 1 and 2 would receive funding at the lowest levels and be assigned probationary statuses. Type 1 providers would be at imminent risk of contract termination, and Type 2 providers would be frequently monitored and at risk of possible contract termination. Type 3 would be considered the minimal performance level that is acceptable and receives base funding levels. Types 4, 5, and 6 would receive progressively more funding (i.e., incentive funding) and would be given preference for treating high-risk/high-need and specialized clients. Later in this report, we address concerns related to OCC adopting a performance-banded system, but first, we provide a deeper dive into PBC and performance payment strategies.

Performance-Based Contracting Payment Models

Performance-based contracting is a tool for governments to better understand the effectiveness of their funded programs, incentivize performance improvements, and gain confidence that public dollars are yielding desired outcomes. A PBC framework relies on developing metrics against which performance can be measured and tying all or some payment to successful demonstration of those metrics. In social service contracts, these metrics can be based on the quality of contract activities and their associated results or outcomes. Performance-based contracting is intended to incentivize service providers to improve performance while making government spending more effective and cost efficient.

Outcome-based contracts are a more rigorous subset of performance-based contracting where payment is partially or wholly based on the demonstration of achievement of predetermined outcomes (Hawkins, Bieretz, and Brown 2019).¹ For example, to reduce recidivism, a government may be interested in contracting with providers that deliver career services to people recently released from prison. In a traditional fee-for-services model, payment would likely be tied to the number of classes provided. In an outcome-based contract framework, payment would be tied to an outcome metric, like the percentage of clients who attained and retained full-time employment after completing a program, or the percentage of clients who did not commit a new crime in the two years after the program.

Throughout this report, we refer to Colorado's current and desired plan as a performance-based contract, although we recognize that it is more specifically an outcome-based contract. There are five main outcome-based contract strategies that governments can use to incentivize outcomes (Hawkins, Bieretz, and Brown 2019): Those strategies are as follows:

- **Unit of outcome achieved.** Governments can assign a value for achieving an outcome and pay on a per person, per outcome basis.
- **Weighted incentives.** Governments can add weighted incentives or performance-measure adjustments to compensate providers for serving hard-to-serve clients or populations.
- **Percent changes in outcomes.** Governments can pay providers based on a percent increase or decrease of an outcome. Governments must first establish baselines against which percent change can be measured.
- **Performance-banded payment models.** Governments assess provider performance and place them within different payment bands such that higher-performing providers are placed in the higher bands and receive greater compensation.

- **Incentive payments.** Governments pay for basic service delivery and incentivize service providers to meet established outcome targets.

Building off these strategies, Urban has identified three potential PBC approaches best aligned with the goals Colorado outlined for its 2015 PBC plan. The first is an **incentive payment** approach. This approach gives providers the opportunity to earn additional funding and/or contract extensions if they meet or exceed established performance targets regarding activities, quality, or outcomes. In this approach, government agrees to pay for the cost of the services provided via a traditional fee-for-service model, but awards additional payments to service providers who exceed previously identified performance targets (Martin and Desenberg 2009). Covering all or most of the providers' cost of service delivery enables them to focus on improving performance without worrying about the consequence of reduced revenue. This approach presents a minor risk for providers and programs. If they do not achieve superior performance, they will not suffer any financial losses, but will need to remain compliant with baseline contract requirements in order to compete for future contracts. And in some incentive models, providers or programs that achieve superior performance can receive additional government work without having to compete for contract renewal (Martin and Desenberg 2009).

BOX 1

Incentive Payment in Pennsylvania

In July 2014, the Pennsylvania Department of Corrections entered into three-year performance-based contracts with its residential Community Corrections Centers and Community Contract Centers. Programs have to stay within a target range of the risk-adjusted recidivism rate (rearrest and reincarceration for parole violations within six months of successful program completion) and if they do, they can earn a modest incentive.

The annually calculated baseline recidivism rates are adjusted based on the risk profile of clients of the program (low, medium, high) based on majority (i.e., 50 percent or more) of the client population's LSI-Revised score being low (0–17), medium (18–27), or high (27–54). The target range for each outcome is within one standard deviation above or below the mean for that risk group. The recidivism rates for each group were as follows:

- low = 4.4–28.4 percent
- medium = 10.4–34.4 percent
- high = 22.7–46.7 percent

In addition to recidivism, other target metrics for additional incentives are: program completion (57.0–78.2 percent), unexcused absence/escape (10 percent), employment (75 percent), program audits (95 percent), security audits (100 percent), and operation audits (100 percent).

Programs exceeding the baseline in the first three months can get a bonus of 0.5 percent in the following contract year for reaching the target range. If a program consistently meets the target, it can earn 1.0 percent in later contract periods. The first time a program does not meet its target, it gets a warning, and the second time, it could be held in contract default. Losing a contract from poor performance is a very real risk to programs and a way the state can use contract renewals to encourage performance improvement.

Source: Commonwealth of Pennsylvania (2013) Invitation for Bid for Housing and Treatment Services, IFB #6100024114.

The second PBC type is a **performance-banded payment model**. Performance banding is a mechanism whereby PBC providers are evaluated and ranked into “bands” or areas of high, mid-range/average, and low performance. The performance bands determine payment. Providers placed in the high-performance band achieve a larger payment than those placed in the lower bands. When the placement into bands is reviewed annually, providers will be incentivized to do well so that they are not placed in lower bands in subsequent years. When using disincentives, providers in the high-performance bands that receive a penalty for an unsuccessful outcome will see a penalty that is lower than that for providers in the mid-range/average and low performance bands. This mechanism gives lower-performing providers an incentive to improve in evaluated dimensions when they otherwise may not be able to compete with the highest-performing providers (Hawkins, Bieretz, and Brown 2019). This structure provides minor risk to providers if penalties are used because they may not fully recoup their costs, and the risk of less payment is greater for those in the lower bands.

BOX 2

Performance-Banded Payment in Tennessee

In 2005, the Tennessee Department of Children’s Services began exploring performance-based contracts to improve the incentives in its payment structure. It hoped to replace its pay-for-service structure, which unintentionally disincentivized achieving permanency for the children in programs’ care. The department launched its PBC initiative in July 2006 with five agencies, and by July 2009, all department-contracted providers were working within the PBC framework. The shift from non-PBC to PBC relied on three key steps: observing variations in provider performance, monitoring provider performance, and changing the reimbursement system to link payments or incentives to performance indicators.

Observing variation, or adjusting for risk, leveled the field for providers serving children with different needs or varying degrees of risk. To monitor performance, the department introduced the Baseline-Target-Actual method to measure changes in provider performance in four priority areas: exits to permanency, care days used/length of stay, rate of reentry to care, and nonpermanent exits. Baselines and targets were set based on child age, adjudication type, and amount of time already in care. The following targets were established for performance improvement:

- Care days: a 10 percent decrease in the baseline number of care days used during the period
- Permanent exits: a 10 percent increase in the baseline number of youth achieving a permanent exit
- Performing within a specified corridor for the number of reentries

The Department of Children’s Services introduced financial incentives and shifted payments from fee-for-service to a model that requires permanency outcomes before payment. When a provider met or exceeded the target that the department established, it reinvested savings with that provider. If a provider performed below the established baseline, it was penalized. In five years, Tennessee’s foster care service providers lowered care days by 235,000 and earned \$20 million in reinvestments as a result of their service improvements.

Sources: “Performance-Based Contracting in Tennessee’s Foster Care System,” Center for State Foster Care and Adoption Data, accessed October 8, 2020, https://fcda.chapinhall.org/wp-content/uploads/2012/10/2012_TN-PBC-case-hx2.pdf; “Performance-Based Contracting in Tennessee’s Foster Care System,” Center for State Child Welfare Data, accessed October 8, 2020, https://fcda.chapinhall.org/wp-content/uploads/2018/02/Performance-Based-Contracting-in-Tennessee_2.2018.pdf.

The third type of PBC that Urban identified that could be relevant to OCC involves **payment by unit of outcome**. Output frameworks, like paying per diem for bed use, link provider compensation to the amount of services provided. Providers are paid a fixed rate for each “unit of service” provided. Similarly, outcome frameworks link payment to accomplishment of individual-level outcomes, and payment is based on the number or units of outcomes achieved. This presents moderate to major risk for providers because accomplishing outcomes takes time and is generally the performance measurement over which providers have the least control. Governments should be extremely careful when setting outcome prices because it is easy to underprice outcomes and shortchange providers, or overprice outcomes and overcharge the government.

Importantly, governments have the option to incorporate more than one of these strategies at once, creating a hybrid model. The most important components to incorporate are (1) creating a concrete link between services and program objectives, (2) evaluating the true cost of services or programs, and (3) tracking and verifying outcomes.

It is crucial that governments and providers work together to determine fair and equitable outcome targets that governments can use to determine payments. One way to do this is to transition gradually to a PBC framework by adopting a **hold harmless period** during which providers are evaluated on their performance but payment is not yet linked to that performance. This gives providers the opportunity to familiarize themselves with the PBC structure and identify improvements. This period also helps governments avoid unfairly penalizing providers while the structure is being refined.

Recommendations for Risk-Adjusted Outcome Measures

One important component of the PBC plan for Colorado's community corrections programs was the development of risk-adjusted outcome measures. The Office of Community Corrections requested that Urban develop a baseline outcome assessment, a method for adjusting those outcomes by risk, and a method for evaluating program performance against those outcomes. Using the results from the stakeholder survey and Community Corrections Information and Billing data, the Urban team identified and generated relevant program outcomes. The final outcome measures are gaining or retaining employment, change in risk score, and termination reason (i.e., successful completion, escape, technical violation). We also used the recidivism data from the courts to generate indicators of conviction for a felony one and two years from program start date. We calculated outcomes at the program level and at the state level over a five-year period of interest (FY 2014–18). For recidivism, we use FY 2014–17.

Urban developed performance targets using the historical data (FY 2014–18) that account for variation in risk levels. We calculate the target for low- and medium-risk programs as the average of all clients in the state on a given outcome because no program has exclusively low- or medium-risk clients. We calculate the target for high- and very high-risk programs as the average for all high- and very high-risk clients in the state. The target range for positive outcomes (i.e., employment gain and successful completion rate) is equal to or greater than the mean. The target range for negative outcomes (i.e., felony conviction, escape, and technical violation) is equal to or less than the mean. For LSI change, the target is equal to or less than the mean. The performance targets for each risk group (low/medium and high/very high) are shown in table 3.

TABLE 3

Risk-Adjusted Outcome Performance Targets

Measure	Low/medium risk	High/very high Risk
<i>Positive outcomes, equal to or greater than:</i>		
Successful completion rate	55%	49%
Employment gain/retain	75%	69%
<i>Negative outcomes, equal to or less than:</i>		
Escape rate	16%	19%
Technical violation rate	27%	30%
Felony conviction two years from program start	17%	19%
Level of Service Inventory change equal to or less than:	-4.6	-5.7

Urban then categorized each program by risk level and calculated whether its performance in FY 2019 (FY 2018 for recidivism) met the target for that risk group. Figures 1 and 2 show the proportion of programs that met each target in FY 2019 by risk group. For successful completion, 43 percent of low- and medium-risk programs and 83 percent of high- and very high-risk programs met the target. For felony reconviction within two years, 25 percent of low- and medium-risk programs met the target, and 68 percent of high- and very high-risk programs met the target. Further information about the data and methods are available in appendix B.

FIGURE 1

Proportion of Programs Meeting Performance Measure Targets in FY 2019

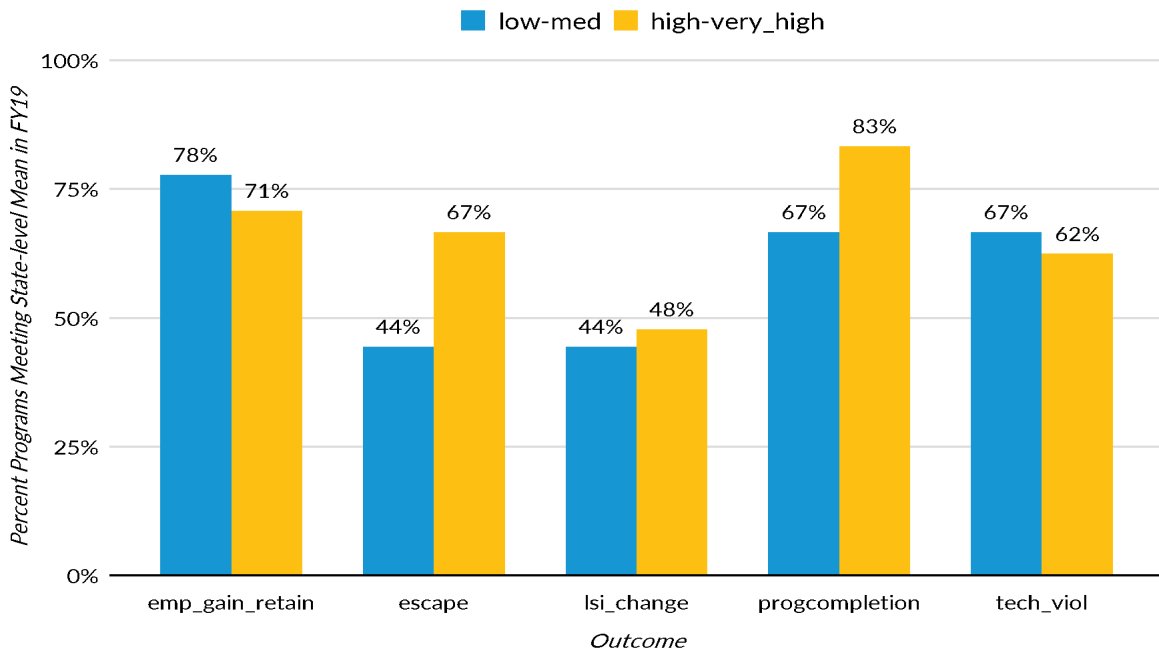
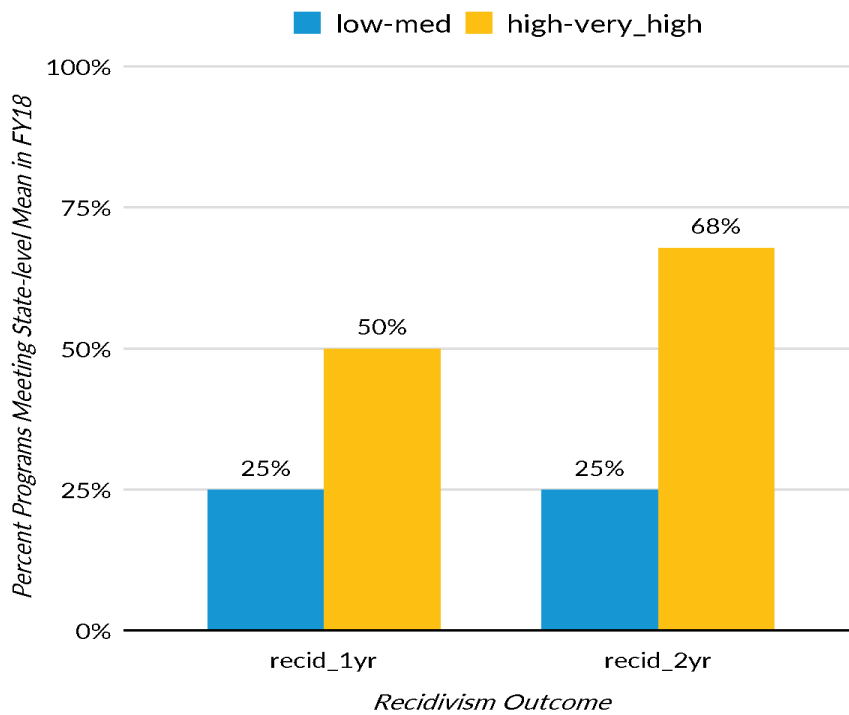


FIGURE 2

Proportion of Programs Meeting Recidivism Targets in FY 2018



Recommendations for New Measures

The survey of community corrections stakeholders revealed several outcome measures that would be beneficial to collect and are not currently collected in OCC's records-management system (i.e., Community Corrections Information and Billing). One measure that was ranked very highly in the survey was treatment and programming that matches risks and needs. In the current system, some risks and needs are captured (e.g., level of substance use need), while receipt of treatment and programming is inconsistently captured across community corrections programs. In the new records management system, more risks and needs that match the available treatment and programming should be measured. Similarly, receipt of treatment and programming should be tracked more consistently across programs.

Another potential measure is staff retention. Most outcome measures focus on client-level outcomes rather than programs' characteristics. However, during interviews and on the survey, many stakeholders indicated that staff retention is a signal of program success because retaining staff allows programs to learn and develop more EBPs. Survey respondents also indicated that client progression through the level system is another way to measure client success. Successful program completion captures successful advancement through all of the levels, but does not capture clients who make it through several levels before being terminated for other reasons. Many programs already track progression through the level system in their records-management systems, and it would be helpful to track the levels in Community Corrections Information and Billing.

Considerations for Risk-Adjusted Outcome Measures

In developing multiple outcome measures, the goal is to support a more holistic measure of program success. However, as with all data measurement, there are limitations and considerations for these measures. First, there is no perfect way to adjust for risk at the program level. No program has exclusively low- or high-risk clients. Based on its review of data, the Urban team determined that the most straightforward way to adjust for risk is by the risk level of the majority of a program's clients. Because programs with a majority of low- and medium-risk clients also have high-risk clients (and vice versa), the performance of programs at low/medium risk and high/very high risk may be more similar than what is observed for the state-level averages.

There are also limitations to the specific outcome measures. Regarding felony reconviction as the recidivism measure, the likelihood of conviction for felony charges may vary throughout the state. For example, clients charged with felonies in one county may face prosecutors and judges with different

propensities for moving cases forward and for convictions than in other counties. However, using felony reconviction rather than any arrests or new charges likely reduces variation across counties. Regarding successful completion, programs may change their practices in unanticipated ways to increase their successful completion rate. For example, programs could keep clients longer in hopes that they progress through the level system and successfully complete rather than terminate them unsuccessfully. Some of these adjustments may be positive and may represent movements toward practices that are more evidence based, but it is important to consider how programs may make unanticipated changes based on the outcomes evaluated in PBC.

Recommendations for PBC Payment Models in Colorado

The State of Colorado wants to move all community corrections contracts from a fee-for-service payment model to a performance-based model that enhances public safety and incentivizes the adoption of EBPs. The current system uses per diem rates for reimbursement with extra supplemental or differential payments for specialty programs like intensive residential treatment. For the purposes of this report, supplemental payments for specialty beds are not included in this PBC payment model. However, we suggest that OCC proceed with its plan to award contracts for specialty beds based on meeting or exceeding the benchmarks of the new performance-based payment system.

Urban reviewed the seven principles from the 2015 PBC plan and used them to identify the criteria that the future PBC model should meet. Those criteria are as follows:

- Adjust for client risk to prevent penalizing providers for working with clients with varying degrees of risk.
- Be transparent.
- Have a nonpunitive financial structure.
- Incentivize improvement in outcomes (as measured by the risk-adjusted outcomes).
- Incentivize use of evidence-based and best practices (as measured by PACE).
- Incentivize public safety (as measured by Core).
- Include metrics under the control of programs (as measured by PACE and Core).
- Transition programs to the new payment structure as smoothly as possible.

In addition, PBC models work best when there is a gradual transition to the new performance-based payment model.

Earlier in this report, we identified three performance-based payment strategies that OCC might be interested in using in its PBC implementation plan. The first was incentive payments, the second was a performance-banded payment model, and the third was payment by unit of outcome. Upon further examination, only the first strategy seems promising. We discuss incentive payments in greater detail and propose a payment model after briefly explaining why payment by unit of outcome is not feasible and after a section detailing issues with OCC using a performance-banded payment model.

Payment by unit of outcome would require recreating the entire payment system so that programs only receive payment after an outcome is achieved. This would require programs to wait until after a client is terminated to receive payment, which does not seem feasible. It would also require developing a way to price each client-level outcome, which would require a lot more data on the actual program costs and any regional variations. The next section explains why a performance-banded payment model is not possible at this point in time.

Issues with a Performance-Banded Payment Model

The goal laid out in the 2015 PBC plan was to combine scores on the risk-informed outcomes with results from PACE and Core to create one composite score, examine the variation in scores, develop performance bands, and then use each program's composite score to determine where it would fit on the performance-banded payment model. However, Urban advises against using this banded payment model because there are issues with the timing of the PACE and Core assessments, with scoring, and with variation that pose challenges to this approach.

When the Tennessee Department of Children's Services decided to use performance-based contracting to improve child welfare outcomes, the first step was to use past performance data to assess the providers and observe variations in outcomes among them.² Urban is doing this for the risk-adjusted outcome measures, but more data are needed to observe meaningful differences in programs' PACE and Core scores and to observe each program's year-over-year outcome changes.

TIMING ISSUES

PACE and Core are time-intensive assessments that involve tape reviews, interviews, observations, and document reviews. As such, it will take two or three years to get all programs assessed by both tools. Because programs will not be assessed each year on the PACE and Core (though the risk-adjusted

outcomes will be calculated annually), it is difficult to determine whether the PACE or Core score for a given year should go into the composite score, especially for a program that may not receive one or the other assessment in a given year or whose most recent assessment is two or three years old.

Baseline assessments of programs on PACE and Core are still in process. Some programs may not have been assessed yet on both tools because the COVID-19 pandemic disrupted the schedule for administering the baseline assessments statewide. Moreover, postpandemic may not be the best time to finish the assessments if OCC is looking to create statewide benchmarks, because programs may have reduced their population, adjusted their practices to meet public health guidance, or have limited access to outside service providers. Also, OCC started PACE in 2017 and has progressed with administering and scoring the assessment. Early scores may not be as reliable as the scores in later years when the process for rating programs on each factor was better established. A longer baseline period with multiple assessments per program may be necessary to establish a baseline that is reliably measured and accounts for disruptions caused by the pandemic.

SCORING ISSUES

The 2015 PBC plan outlined a process to calculate a composite score that combines the scores from PACE, Core, and the outcome measures. The scales for each assessment are different: Core scores on a scale of zero to three, and PACE scores on a scale of zero to four. It is therefore unclear whether the raw scores are appropriate or whether they should be weighted for alignment. For example, a score of three on Core indicates that a program “meets expectations,” whereas a score of three on PACE indicates that a program “excels at implementation of EBPs and demonstrates advanced competency.” These scales are not directly comparable and it is unclear how to adapt them to fit a potential composite score.

Furthermore, the PACE and Core baseline scores so far are low, with most programs scoring lower than a two. Any composite number would also be very low until improvement is shown in the scores. Without more years of data, it is difficult to know where to set the targets for improvement and to create a reliable method to stratify the scores. It is essential to know that the PACE and Core targets are realistic to assess whether composite scores would be punitive over time.

Moreover, the measure with the largest weight in the composite score would likely be the primary measure that programs focus on improving. Accordingly, expected improvements in the weighted, composite measure should be realistic and align with Colorado’s highest priorities.

ISSUES WITH VARIATION IN THE DATA

There are also issues related to the variation in data and the ability to develop a meaningful stratification scheme for the banded payment model. With little variation in scores across programs and a small range of values for the scales, it is difficult to determine bands that meaningfully distinguish between performance levels. Additional evidence or analyses would be necessary to determine the link between differences in composite scores and appropriate performance-payment bands. PACE and Core are new assessment tools, so there are no historical data to predict changes in scores. To develop effective stratification, it would be ideal to know the probability of scores changing before setting up the bands. This would help prevent against unintentionally penalizing programs and would allow for the selection of targets that incentivize change. Given that the scores are still in baseline measurement and that they are very similar across programs, Urban recommends against using a performance-banded PBC model.

Incentive Payments

When reviewing the work of other states and other performance-based contracts, Urban identified one key payment strategy that OCC could use to structure its PBC payment model: **incentive payments**. This approach includes benchmarks that are risk-adjusted outcome targets, which are lower for clients who are harder to serve and are at higher risk. Urban recommends that Colorado adopt a PBC payment model that covers the next five years and couples a baseline per diem payment with incentives earned by meeting state-level risk-adjusted outcome targets and demonstrating improvement in PACE and Core scores. Consistent with the recommendation for a hold harmless period, Urban recommends that in the initial years of the PBC model, the per diem is provided at 100 percent of the baseline. These “automatic” per diem payments would then be reduced gradually as incentive payments tied to performance are introduced. It is important to note that even after transitioning to the performance-based payment system (i.e., year five), the per diem payments will make up the vast majority of payment to providers with the performance payments providing an additional incentive.

How Would the Payment Model Work?

The per diem rate for daily bed use would presumably be set by the Colorado Department of Public Safety and approved by Colorado’s Joint Budget Committee. Each incentive is a portion of the state-approved per diem rate. Over time, the baseline amount for the per diem that is earned (not tied to performance) is reduced and the incentive payment (tied to performance) is increased. Because the

PACE and Core scores measure compliance with EBPs and the Community Corrections Standards (which are considered to be operational requirements), the model assumes that the only incentives that can be earned *above the baseline* per diem rate would be purely from improved outcomes. The maximum “bonus” that providers can receive above the baseline per diem is relatively low for two reasons: (1) evidence from Pennsylvania suggests that an incentive of one percent coupled with the potential loss of contract were effective at incentivizing provider behavior, and (2) to ensure a model funded with taxpayer dollars is cost efficient, the model should seek the lowest possible payment that is effective at incentivizing performance improvement.³

The model uses **risk-adjusted outcome targets** so that programs are not penalized for serving higher-risk clients. The annually calculated risk-adjusted outcome targets will be based on the five-year historical averages of clients across the state. Program risk is calculated based on the risk level of the majority (i.e., 50 percent or more) of the program’s client population for the year. We recommend using one positive outcome (successful completion) and one negative outcome (felony reconviction within two years of program start) in the model. The recidivism and successful completion outcomes will be calculated using data that are at least two years old to determine whether the outcome was realized. For example, the FY 2022 recidivism outcome will be based on clients released in FY 2020. Each year, programs’ outcome measures can be assessed on their program-level average as it compares with the average for all clients in the state. Higher-risk programs will need to reach a lower outcome target for positive outcomes than lower-risk programs. Conversely, higher-risk programs will meet a higher target on negative outcomes than lower-risk programs.

Programs can also receive an incentive payment for improving their scores on the **PACE and Core assessments** as compared either with their previous scores or a state-calculated average. Urban recommends the former, because there are insufficient data to know the variability in scores and the expected improvement on each assessment over time. The model assumes that PACE and Core assessments will continue on a regular schedule, and because the administration cycles may not align perfectly, in early years, the programs can earn the incentive by showing improvement in either of the two assessments.

Another incentive for improved performance that can be incorporated into this model is determining which programs get **future specialty bed contracts or contract renewals**. The Office of Community Corrections can also develop a warning system whereby if a program does not show improvement over a certain number of consecutive years, the judicial district can cancel its contract. Conversely, OCC can be explicit in the PBC model that a program’s performance on risk-adjusted

outcomes, PACE, and Core will be used to consider contract renewals with the judicial districts and specialty bed contracts.

To better understand the models, Urban presents the following example of the gradual transition model in practice. Assuming that the FY 2024 per diem is \$50.00 per day for a regular community corrections bed, every program would receive 98 percent of \$50.00, or \$49.00 for every client served regardless of their performance on the metrics. If the program met its performance targets on either their PACE or Core score (assuming it received one of the two assessments each year), it would get another \$1.00 per client per day. If the program met or exceeded the target for successful completion, it would receive an additional \$0.50 per client per day. Similarly, if it performed well on the recidivism metric, it would get another \$0.50 per client per day. This makes \$51.00 the maximum payment for the gradual model in FY 2024, provided the program did well on the measures for all three metrics (i.e., PACE/Core, successful completion, and recidivism). If a program did well on PACE or Core and on one of the outcome metrics, it would get \$50.50. If the risk-adjusted outcome metrics for FY 2024 are based on previous years' data (e.g., FY 2022–23), then the payment for reaching the outcome targets could fall in the second half of FY 2024 so that the base per diem and incentive payments can be paid in the same year.

Benefits of the Models

Urban has identified the following benefits of these proposed models:

- **Gradual adjustments.** The tables include a gradual reduction in the per diem, and in the first year, programs will get full payment. This should allow programs to become familiar with the performance metrics and focus on taking needed steps to improve their scores without risk of reduced funding.
- **Annual adjustments.** The state budget committee and OCC can increase or decrease the per diem rate based on the current state budget and fiscal environment.
- **Flexible targets.** The rate table offers a flexible structure where OCC can assess programs' performance on risk-adjusted outcomes as compared with the state average and programs' year-over-year improvement on PACE and Core.
- **Flexible schedule.** Allowing programs to earn the incentive by improving on either the PACE or Core assessment affords the state flexibility for changing how often each assessment is administered.

- **Minimal contract changes.** This new payment structure can be incorporated in judicial district and program contracts by simply amending the letter that includes the fiscal year’s bed allocation and per diem amount because it still relies on a per diem payment structure.

TABLE 4

Gradual Transition Model

	FY 22 (7/1/21– 6/30/22)	FY 23 (7/1/22– 6/30/23)	FY 24 (7/1/23– 6/30/24)	FY 25 (7/1/24– 6/30/25)	FY 26 (7/1/25– 6/30/26)
Base per diem	100%	99%	98%	97%	97%
Completion	baselines	1%*	1%*	1%*	1%*
Recidivism	baselines	1%*	1%*	1%*	1%*
PACE or Core	baselines	baselines	2%	3%	3%
Max payment	100%	101%	102%	102%	102%

Note: * payment based on data from previous one to two fiscal years.

TABLE 5

Progressive Risk and Reward Model

	FY 22 (7/1/21– 6/30/22)	FY 23 (7/1/22– 6/30/23)	FY 24 (7/1/23– 6/30/24)	FY 25 (7/1/24– 6/30/25)	FY 26 (7/1/25– 6/30/26)
Base per diem	100%	98%	96%	94%	92%
Completion	baselines	2%*	2%*	3%*	4%*
Recidivism	baselines	2%*	2%*	3%*	4%*
PACE or Core	baselines	baselines	4%	6%	8%
Max payment	100%	102%	104%	106%	108%

Note: * payment based on data from previous one to two fiscal years.

Decisions to Be Made

Urban has presented two potential models for a performance-based incentive payment model. The first step is to **select the gradual or the progressive model** based on OCC’s priorities and the state fiscal environment. Another important consideration is what level of baseline per diem is needed to ensure that clients receive quality treatment from community corrections programs. If Colorado adopts the progressive risk and reward model, it should commit to keeping the baseline per diem rate steady and avoiding further reductions from the baseline amount.

Another way to discourage poor performance is to create a nonmonetary disincentive related to poor performance, which can be useful when determining which programs get future specialty bed contracts or contract renewals. To do this, OCC can **develop a reward and warning system** whereby if programs do not show improvement in a given number of consecutive years, OCC can cancel their

contracts. Conversely, OCC can be explicit in the PBC model that performance on outcomes, PACE, and Core will be used to consider specialty bed contracts.

The above model only includes rates for the percentage of the per diem assigned to meeting targets on outcomes and performance on PACE or Core. The Office of Community Corrections will still need to **determine the target score comparisons for PACE and Core**. Urban has developed risk-adjusted outcome calculations that OCC can use to create annual outcome targets, but OCC will need to decide whether improvement on PACE and Core will be calculated against programs' past scores or against a state average.

Outcomes can be measured annually and in theory would be positively impacted by using the evidence-based practices and by meeting the requirements of the 2017 Community Corrections Standards. The Office of Community Corrections plans to administer each assessment annually, with half the programs receiving Core and the other half receiving PACE, so that programs get either a PACE or Core assessment every year. Programs' current PACE and Core scores are low, and although that leaves room for improvement, observable changes in performance do not happen overnight. The Office of Community Corrections will need to determine the **frequency of calculating benchmarks for success on PACE and Core** based on its capacity to administer the test frequently. This decision can be revisited as more data are collected on score changes over time and as OCC gets a better idea of when improvements could be observable in future assessments.

Lastly, Urban has created six risk-adjusted outcome measures: successful program completion, gaining or maintaining employment, LSI score reduction, termination for technical violations, termination for escapes, and felony reconviction. These six outcomes are described in the section on recommendations for risk-adjusted outcome measures and in the technical appendix. Urban suggests that OCC **select the two outcomes of greatest importance** and use their risk-adjusted outcome measurements as state-level targets for the high- and low-risk programs in the state. In tables 4 and 5 above, Urban proposed using successful program completion and recidivism. Choosing to use all six outcomes would make it more difficult to calculate programs' payments in the PBC model because it would reduce the percentage incentive allocated to each outcome to a fraction of a percent. Moreover, even though the outcomes are interrelated, if all the outcomes are equally weighted, it is difficult for programs to know which outcomes to focus on to improve their performance. However, even though Urban recommends using only two outcomes that are tied to payments, Urban recommends that OCC track all six outcomes.

Recommendations for the PBC Implementation Plan

To implement Urban's proposed PBC plan, Colorado will need to consider adjustments to data systems, training, future PACE and Core administration, and the payment structure. In this section, we describe these steps in greater detail.

Data Systems

The first step in implementing a performance-based contract is ensuring that the data that will be used to evaluate performance are collected in a complete, accurate, consistent, and timely manner to ensure outcome measures are reliable and suitable to be used for payment. The Office of Community Corrections is in the process of creating a new data system for programs to track their clients. It needs to ensure that the new system includes ways to track and measure outcomes of interest to OCC and the state, such as collecting additional data on the types and dosage of treatment that clients are receiving. After the system is proven to be sufficient for tracking, OCC should work with programs to improve data entry. It will also need to ensure that it has adequate staffing capacity to administer PACE and Core, and calculate the risk-adjusted outcomes. Lastly, OCC should develop an annual schedule for linking and processing the administrative data that will be used to determine outcome scores for each program participating in PBC.

Training

Because PBC will be a new payment model for judicial districts and their programs, Urban suggests that OCC create a training around earning the incentives in the payment model. The training should include information on how the risk-adjusted outcome measures are calculated, should be delivered to the judicial districts and programs before the transition to the PBC payment model, and should include the schedule for outcome calculation so programs know when payments will be made for incentives. This training could also include a sample of the new annual allocation letters that are incorporated into the contracts.

After the baseline outcome measure scores are completed, OCC leadership should meet with each program to discuss the baseline results and share their historical outcome data. The Office of Community Corrections should provide suggestions for how programs can improve their scores in the next year. This meeting could continue annually as a debrief call on how each program scored in annual outcome calculations.

Future Core and PACE Administration

Urban suggests that OCC try to simplify the PACE and Core assessment tools. At this time, it is difficult to identify specific factors or items within each tool to modify because the baseline assessments for all programs are not complete and the tools have not been validated. However, Urban proposes several ways OCC could modify Core and PACE and administer them differently going forward.

Urban conducted reliability analyses on Core and PACE to assess the consistency of the factors and domains of each assessment. More information on the specific analyses and results is available in the technical appendix. Although these analyses are preliminary because not all programs have received a Core or PACE assessment, they are helpful for examining which factors may be more promising than others. Overall, the seven factors of PACE load well together, and Factors 2, 3, and 6 (motivation, target intervention, and violation response) show promise. For Core, the 14 standards may load well together once scores exist for every program.

Once OCC completes PACE and Core for every program, these reliability and internal validity analyses should be repeated to better identify specific domains and items that have more validity and those that should be removed or modified. With each wave of these assessments, the analyses should be repeated so that OCC can establish that these assessment tools are valid for Colorado's community corrections programs.

One way to more simply administer the assessments is to assess fewer factors or standards. After the validity and reliability analyses are complete on the baseline assessments, OCC could identify which factors or items do not show promising results and remove them from the assessments.

Moreover, OCC could break up the administration of Core and PACE so that the assessments on some of the factors are measured during each performance cycle and others less frequently. During interviews with program stakeholders, many emphasized that their programs could not realistically improve on every evidence-based practice measured in PACE in a given year because it takes time to train staff and shift practices. One option to simplify PACE would be to use the initial full assessment as a baseline and then select two or three factors for evaluation during the next performance cycle. In this way, programs would have a clear idea of which EBPs to improve on, and OCC could provide more focused technical assistance in those areas.

Lastly, to increase the transparency and simplicity of the measurement method, the factors or standards could be modified to include as many binary, objective checkboxes as possible. One

possibility would be to use some of the administrative data outcomes that are calculated as parts of Core and PACE, especially compliance-based metrics that can be measured each year.

Transitioning to the New Payment Model

If Colorado is concerned about having funding to make future incentive payments in excess of baseline payments, it could budget based on the full per diem amount and create a sinking fund that invests the money that might be reasonably expected to be saved by DOC when OCC reduces terminations and recidivism events that result in returns to prison. The fund could also include deposits of money budgeted, but not spent, in years that programs do not meet their benchmarks. This sinking fund could be used for incentive payments in future years. If the fund grew, the state could consider allowing OCC to increase the incentive payments related to achieving good outcomes.

The performance-based payment model will only work if it has the support of the state leadership and is fully funded. The Office of Community Corrections should work with relevant state and local stakeholders to achieve buy-in for PBC and work with the state to approve whatever PBC model is chosen. Urban acknowledges that legislative changes may also be needed to pay performance-based incentives, but those changes are outside the scope of this report. Lastly, since the models rely heavily on the per diem rate, the Joint Budget Committee will need to work with the Department of Public Safety to set an appropriate per diem rate per the state budget.

Overall, the current judicial district contracts and subcontracts with programs could continue to be used as long as the funding allocation letters include the per diem rates for the fiscal year and the incentive structure for that period. The Office of Community Corrections should also develop a schedule for calculating outcomes that is clear about what point in time outcomes will be measured and when the associated payments will be received. It and the judicial districts need to be ready to adhere to the payment schedule to adequately reward the programs.

Conclusion

Adopting one of Urban's proposed performance-based contracting models creates an opportunity to enhance public safety by improving outcomes for community corrections clients. The proposed payment models reward programs for superior performance and encourage adoption of evidence-based practices and compliance with security standards. Both models include full reimbursement in year one to incorporate a hold harmless period, which is an important component for ensuring a fair and

transparent transition from a fee-for-service model to one that is performance based. The models afford the state the flexibility to work with OCC to set annual per diem rates based on need and the current fiscal environment. As a next step, Colorado should select one of the two models and proceed carefully with developing a more detailed implementation plan based off the guidance provided in this report.

Appendix A. Technical Appendix

This appendix documents the technical steps that support the report *Performance-Based Contracting for Colorado Community Corrections*. In the report, we use data from the Colorado Department of Public Safety, Office of Community Corrections (OCC) to develop a method for calculating risk-adjusted outcome measures in preparation for implementation of performance-based contracting (PBC) in Colorado’s residential corrections programs. In this appendix, we detail the data sources and methodology used in our analysis.

Background

In 2015, the Colorado Governor’s Community Corrections Advisory Council completed *Performance Based Contracting (PBC) for Colorado Community Corrections: General Attributes of a PBC Plan*, highlighting three areas of performance: safety and security, the extent and quality of evidence-based practices (EBPs) and principles, and risk-informed outcomes. In 2017, the corresponding Colorado Community Corrections Standards, which align these areas of performance with evidence-based best practices in the field, were completed, and the Program Assessment for Correctional Excellence (PACE) and Core Security Audit (Core) processes were finalized. In 2019, the Urban Institute was contracted to consult on the establishment of risk-informed outcomes, and to provide policy and practice recommendations for the implementation of a PBC plan.

Practitioner Stakeholder Survey

To better understand relevant measures of program performance and current program review processes, the Urban research team developed a short anonymous survey for community corrections stakeholders. The voluntary seven-question survey was administered through the online platform Qualtrics and covered questions on useful measures of program performance, current performance review and audit processes, and PBC training preferences. The survey was sent out to 311 staff members via e-mail, and Urban researchers received 83 complete survey responses (a 27 percent response rate). Of those who responded to the survey, 71 percent of respondents self-identified their role as one of the survey response options, and 31 percent selected the “other” category and filled in their role in a text field. The full breakdown of respondent positions can be found in table A.1.

TABLE A.1

Respondent Position or Role within Community Corrections

	Percentage of Respondents
Survey responses (by position)	
Program administrator	17
Program security staff	13
Program case manager	14
Community corrections board member/staff/chair	21
Governor's advisory council member	3
Community-based service provider	3
Other responses (by position)	
Program supervisors and staff	10
Department of Corrections staff	3
Parole or probation affiliated	8
Administrators	4
Uncategorized	6

Note: Numbers do not add to 100 due to rounding.

To help understand the relevant program outcomes to be included in the PBC plan, respondents were asked to provide input on the level of utility particular outcomes have in measuring program performance. Respondents were asked to rate specific outcomes on a scale of one to five, where one indicated a strong disagreement that an outcome was a useful measure of program performance and five indicated a strong agreement that an outcome was useful. The three highest-rated outcomes were participation in services and treatment that match risks/needs (average rating of 4.6), changes in risk scores (average rating of 4.3), and successful program completion (average rating of 4.3). The full list of average ratings for outcomes is included in table A.2.

TABLE A.2

Respondent Ratings of How Useful Outcomes Are in Measuring Program Performance

Outcome	Average rating
Participation in services and treatment that match risks/needs	4.6
Changes in risk scores	4.3
Successful completion	4.3
Staff retention	4.1
Progression through level system	4.1
Employment	4.1
Reincarceration	3.6
Reconviction	3.6
Educational attainment	3.6
Escapes	3.6
Technical violations	3.5
Earnings	3.3

Notes: Responses were measured on a scale of 1 to 5 (1 = Strongly Disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree).

To gain a better knowledge of the level of understanding of the current community corrections performance review and audit system and the plan to move to PBC, respondents were asked to rate their level of agreement with a set of statements on a scale of one to five, where one indicated a strong disagreement with the statement and five indicated a strong agreement. Respondents were also asked whether the PACE and Core security audit process adequately scored the outcomes they were interested in. The average ratings for each statement are included in table 3. On average, respondents were either neutral or agreed with the statements posed, with average overall ratings between three and four. The statement which received the lowest average rating of 2.9 regarded whether PACE adequately scores a facility’s use of EBP. The average ratings can be found in Table A.3.

TABLE A.3

Respondent Level of Agreement with Statements about Current Performance Review and Audit Systems

Outcome	Average rating
I understand the current community corrections performance review or audit system	3.4
I understand the PACE and how it measures performance	3.2
I understand the Core Security Audit and how it is used to measure performance	3.4
I understand the state’s plan for moving to PBC in community corrections	3.4
The PACE adequately scores your facility’s use of Evidence-Based Practices (EBP)	2.9
The Core Security Audit adequately scores your facility’s level of safety and security	3.0

Notes: Responses were measured on a scale of 1-5 (1 = Strongly Disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree).

The Urban research team utilized the findings from the stakeholder survey to identify program outcomes to include in PBC benchmarks. A number of highly rated outcomes such as successful completion and changes in risk score are included in the final PBC model. In addition, the survey results provided insight into the level of understanding staff have of current review and audit systems and highlighted some challenges the Office of Community Corrections (OCC) may face in implementing PBC.

Data Sources

The data used in our analysis come from several sources. Client-level program data come from OCC’s Community Corrections Information and Billing (CCIB) system. Recidivism data come from the Office of Research and Statistics and the Colorado Judicial Department. PACE and Core data for each program also come from OCC.

Community Corrections Information and Billing

The client-level data cover all releases from community corrections programs between 2008 and 2019 and were pulled directly from the CCIB system. Each observation in the dataset represents a unique treatment period for a person. The data received from OCC include client-release information for all the community corrections programs across Colorado. Data from CCIB contain a large number of variables about clients including demographic information, intake assessment outcomes, levels and types of treatment received while in a program, and program termination reason, among other variables. Client release data were delivered in CSV format and were broken down by type of stay (e.g., Cognitive Behavioral Therapy, Intensive Residential Treatment [IRT], Dual Diagnosis, Non-Residential, and Residential). We use the CCIB Cognitive Behavioral Therapy, Dual Diagnosis, and Residential data to develop baseline outcome measures at the program level. The process for identifying and developing these outcome measures is further described below.

Recidivism Data

The data related to client recidivism were retrieved from the Office of Research and Statistics via an agreement with the Colorado Judicial Department. The recidivism data provided cover all new court filings for community corrections clients released between 2014 and 2019. We use these data to generate indicators of felony convictions one and two years from program start date and program release date.

PACE and Core Security Audits

Urban researchers also received PACE and Core results from OCC. The PACE tool was developed in 2017 to measure how well community corrections programs in Colorado adhered to the Colorado Community Corrections Standards, which incorporate the National Institute of Corrections Principles of Effective Intervention.⁴ PACE scores programs on seven factors, each with their own subcomponents. The factors are on a scale from zero to four, with zero indicating that the program needs to initiate implementation of EBPs related to that factor and four indicating that program demonstrates full mastery of current EBPs. The Office of Community Corrections is currently in the process of establishing baseline ratings for each program and Urban received baseline PACE data for 29 programs. The average scores for each factor and the overall for the PACE are provided in table A.4.

TABLE A.4
Average PACE Scores

Measure	Average	N
Factor 1 - Risk/Need	1.10	29
Factor 2 - Motivation	1.55	29
Factor 3 - Target Intervention	1.80	29
Factor 4 - Skill Train	0.91	29
Factor 5 - Positive Reinforcement	1.60	29
Factor 6 - Violation Response	1.85	29
Factor 7 - Ongoing Support	1.98	29
Total Score	1.55	29

The Core process was updated in 2017 and measures program compliance with security standards. The Core process rates programs on a number of standards and includes an in-person field visit, client case file reviews, live program observations and camera reviews, and client and staff interviews. The Office of Community Corrections is currently in the process of performing the updated audits on all programs, and as a result, Urban received Core scores for 24 programs with completed scores. The standards are scored on a scale of zero to three, with zero indicating that immediate action is needed on that standard and three indicating that the program meets expectations related to those standards. The average scores for each standard in Core are provided in table A.5.

TABLE A.5
Average Core Security Audit Scores

Measure	Average	N
CS-010 Random Headcounts and Facility Walkthroughs	2.18	24
CS-011 On-Grounds Surveillance	2.40	24
CS-030 Contraband, CS-020 Client Property	1.78	24
CS-040 Random Off-Site Monitoring	1.66	24
CS-041 Furlough Monitors	1.94	20
CS-042 Job Search	1.06	20
CS-044 Home Visits	1.00	21
CS-050 Recording Authorized Absences	2.09	24
CS-060 Substance Testing Processes	2.07	24
CS-061 Entry Urine Sample	2.25	24
CS-063 Confirming Positive Test Results	1.67	21
CS-080 Medications	1.56	24
CS-090 Escape	1.65	24
OMA-020 Milieu Management	1.87	24

Methodology

The research team took several analytic steps to create a PBC implementation plan for Colorado community corrections. First, using the results from the stakeholder survey and CCIB data, we identified and generated relevant program outcomes. The final outcomes included: employment status at program exit, change in risk score, and termination reason (i.e., successful completion, escapes, and technical violation). We also used the recidivism file to generate indicators of conviction for a felony one and two years from program start. We calculated outcomes at the program level and at the state level over a five-year period of interest (FY 2014–18). We calculated recidivism outcomes using FY 2014–17.

We then categorized programs by risk level and calculated target ranges for outcome measures by the risk level of a program. Target ranges for positive outcomes (i.e., employment and successful completion rate) are higher than the mean. Target ranges for negative outcomes (i.e., felony conviction, escape, and technical violation) are lower than the mean. The target range for Level of Service Inventory (LSI) change is lower than the mean, as the average change is negative (i.e., the average score decreases).

Outcomes

We utilized the data outlined above to generate seven outcomes to be included as benchmarks in the PBC plan. We first generated binary variables for each outcome (except for change in risk score, which is calculated as the difference between two scores) at the client level, which were then used to create rates for each outcome at the state and program levels. The process and definitions for creating these outcomes are outlined below.

Sample

The sample included everyone released from a community corrections program between 2008 and 2019 except people who fall under the exclusionary criteria outlined below. Before analysis, we imported the client-level CCIB data into statistical software and appended each dataset to create the full file of clients. In Colorado, the fiscal year runs from July 1 to June 30 of the next year. Because OCC contracts with providers will be based around this annual schedule, we generated a fiscal year variable based off of the service end date of each person and used this variable throughout our analysis. For example, all of the baseline outcome values are generated from release data from FY 2014 to FY 2018. The recidivism outcomes are generated from releases from FY 2014 to FY 2017.

Exclusionary Criteria

In order to generate accurate program outcomes, we excluded a select subset of observations. We excluded the IRT file from analysis because this is a short-term, 90-day program, and people who completed IRT often transitioned to residential programming. A number of the outcomes, such as LSI score change and change in employment, are not fully captured in the data for IRT clients, so we excluded the stay during IRT to ensure that outcomes accurately reflected the greater program population. Since many people subsequently transferred to residential treatment following IRT, they are likely still represented in the final dataset.

In addition to IRT, we also excluded clients with a number of termination reasons that would skew the outcome data if included. The excluded termination reasons include: reject after accept, continuous stay, outstanding warrant/pending crime, transfer to other community corrections program, transfer to IRT, transfer to Short Term Intensive Residential Remediation Treatment, and other.⁵ We excluded these termination reasons because clients who are rejected or transferred do not stay in the programs very long. For clients who terminated with a continuous stay, they likely are still represented in the data set for the next stay. We also excluded clients that terminated for crimes that occurred prior to their entry to the program. In total, 19,944 observations were dropped because of the exclusionary criteria, leaving 63,821 observations in the final dataset of clients released from FY 2009 to FY 2020.

Outcome Definitions

EMPLOYMENT RETENTION OR GAIN

We generated employment outcomes in relation to the population considered eligible for employment. This included people listed as full-time employed, part-time employed, or unemployed and excluded those listed as unemployed because of disability and those with missing values. A person is included as having *retained employment* (0 = did not stay employed; 1 = stayed employed) if they were employed at the time of program entry and at the time of program exit. A person is included as having *gained employment* (0 = did not gain employment; 1 = gained employment) if they were unemployed at program entry, but employed at exit. The outcome of interest was defined as whether or not someone *gained or retained employment* (0 = lost/remained unemployed; 1 = gained or retained employment).

RISK REDUCTION

Risk reduction is defined as the change in Level of Service Inventory score between the initial screening performed at intake and the six-month update. The LSI is a 54-item risk assessment tool that provides a

measure of risk for recidivism and profiles the areas of need that contribute to a high score. A higher score on the LSI corresponds with a higher risk of recidivism, meaning a reduction in the score is viewed as a reduction in the level of risk of a person. Office of Community Corrections programs are instructed to administer an LSI to a client within 10 days of program entry if they have not received an LSI in the six months prior to entry. The LSI is then scheduled to be administered every six months and when a significant event occurs in the client's placement. We utilized the initial and the six-month updates to calculate the LSI change. If a six-month update is not available for the client, the LSI change is not calculated.

SUCCESSFUL COMPLETION

The successful completion outcome is calculated from the termination reason variable provided by OCC in the CCIB data. This variable is coded as one if a person's reason for termination is a successful completion and zero if they were terminated for a technical violation, escape, or a new crime.

TECHNICAL VIOLATION

The technical violation outcome is also calculated as a binary variable (1 = terminated for a technical violation; 0 = terminated for another reason) from the termination reason variable provided in the CCIB data.

ESCAPE

Similarly, the escape outcome is calculated as a binary variable where one indicates that a person was terminated for escaping and zero indicates they were terminated for another reason.

RECONVICTION

Urban received recidivism information for all clients released between 2014 and 2019. These data included new court filings and conviction information at the client (uniquely identified by an offender ID) and residential stay (uniquely identified by a service ID) levels. This meant that the recidivism information matched directly back to a specific residential stay in the CCIB data so that people with multiple stays over time are represented in the recidivism data. That is, we measured recidivism during and after each residential stay that a client may have. For the purposes of this study, we focused specifically on felony convictions. We generated indicator variables for felony convictions one and two years from the start date of community corrections. To generate these outcomes, we merged the recidivism file with the CCIB file using the combination of a unique offender identification number and service identification number. We then generated the indicator variables as a binary where one

indicates a felony conviction during the specified time period from the start date and zero indicates that there was not a recorded felony conviction under those parameters. For this measure, the new felony offense has to occur after the start date. We do not capture felony convictions for offenses that occurred before the start of community corrections.

Level of Analysis

We look at the outcomes defined above at two units of analysis to help us generate PBC benchmarks. In this section, we describe the two levels and also provide tables of the averages for each outcome.

State Level

The first level of analysis was the average for all community corrections clients in the state. For each fiscal year, we calculated the average (mean) of each outcome for all clients released from community corrections programs in Colorado. For example, if 1,000 people were released in a given year and 500 of them successfully completed, the successful completion rate would be 50 percent. We then used this state-level average to set benchmarks for program performance. The mean (average) for each outcome is shown in table A.6.

TABLE A.6
State-Level Average for Outcome Measures (FY 2014–18)

Measure	Average	N
Felony conviction 1yr from start	13%	26,584
Felony conviction 2yr from start	17%	26,584
Employment gain/retain	73%	25,727
Escape rate	17%	26,553
Change in LSI score	-4.62	12,771
Successful completion rate	54%	26,553
Technical violation rate	27%	26,553

Note: Felony reconviction only uses FY 2014–17.

Program Level

The second level of analysis we examined was at the program, or facility, level. The program-level outcome rates are generated by calculating the mean for each program in a given fiscal year and then averaging across all of the programs. For example, if there are two programs and one has a successful

completion rate of 60 percent in a given year and the other has a rate of 40 percent, the program-level average for the two programs would be 50 percent. The mean (average) for each outcome across the programs in the state is shown in table A.7.

TABLE A.7
Program-Level Average for Outcome Measures (FY 2014–18)

Measure	Average	N
Felony conviction 1yr from start	12%	165
Felony conviction 2yr from start	17%	165
Employment gain/retain	73%	165
Escape rate	17%	165
Change in LSI score	-5.32	165
Successful completion rate	55%	165
Technical violation rate	26%	165

Note: Felony reconviction only uses FY 2014–17.

Considerations for Each Analysis Level

The average of each outcome for the state-level and program-level outcomes are very similar. In both methods, the average rate of gaining or maintaining employment is 73 percent across FY 2014–18, and the successful completion rate only differs by one percentage point (54 percent versus 55 percent). However, the unit of analysis for calculating the outcomes impacts the possible methods for adjusting the outcomes by risk, which is discussed in the next section.

Risk Adjustment

Many factors can influence both a client’s success in community corrections and a program’s overall success. Risk assessment tools, such as the LSI, consider many of those factors, such as criminal history, education, employment, relationships, and issues with alcohol and drug use. Scores on risk assessment tools are designed to be associated with risk of recidivism, but the scores are also often related to other outcomes, such as successful program completion, employment, escape, and technical violations. Because individual risk factors are associated with subsequent outcomes, it is important to adjust those outcomes by risk. Similarly, programs that have clients with higher risk scores might have different outcome rates than those programs that have clients with lower risk scores. For this reason, program-level outcomes should also be adjusted by risk.

Risk-Adjusted Performance Targets

A first step is to examine whether the outcomes described above vary by risk in Colorado. Using the state-level method described above, we distinguished whether clients were low or medium risk (defined as an initial LSI score less than or equal to 28) or whether they were high or very high risk (defined as an initial LSI score of 29 or higher).⁶ We then calculated the mean of each outcome for the low/medium and high/very high risk groups separately for all clients released in FY 2014–18 (FY 2014–17 for recidivism). The state averages in figure A.1 and figure A.3 show that the low/medium risk group has higher rates of the positive outcomes and lower rates of the negative outcomes. That is, low- and medium-risk clients have an average successful completion rate of 65 percent, compared with 49 percent for high- and very high-risk clients. The averages for each risk group are significantly different at an alpha level of 0.05 using a chi-squared test.

Another way to examine risk is at the program level. Across the state, the average LSI score for the programs is 30, meaning that most programs on average have high-risk clients. We calculated the average LSI score for each program from FY 2014–18, and then categorized them into low/medium and high/very high using the same cut points as above. We then calculated the average of each outcome for the low/medium risk program and the high/very high risk programs separately (shown as the program averages in figures A.1, A.2, and A.3). Using this method, there is very little difference between the low/medium and high/very high risk programs. For example, their successful completion and escape rates are nearly identical. None of the program-level averages for each risk group are significantly different.

FIGURE A.1

Average Positive Outcomes by Risk Level (FY 14-18)

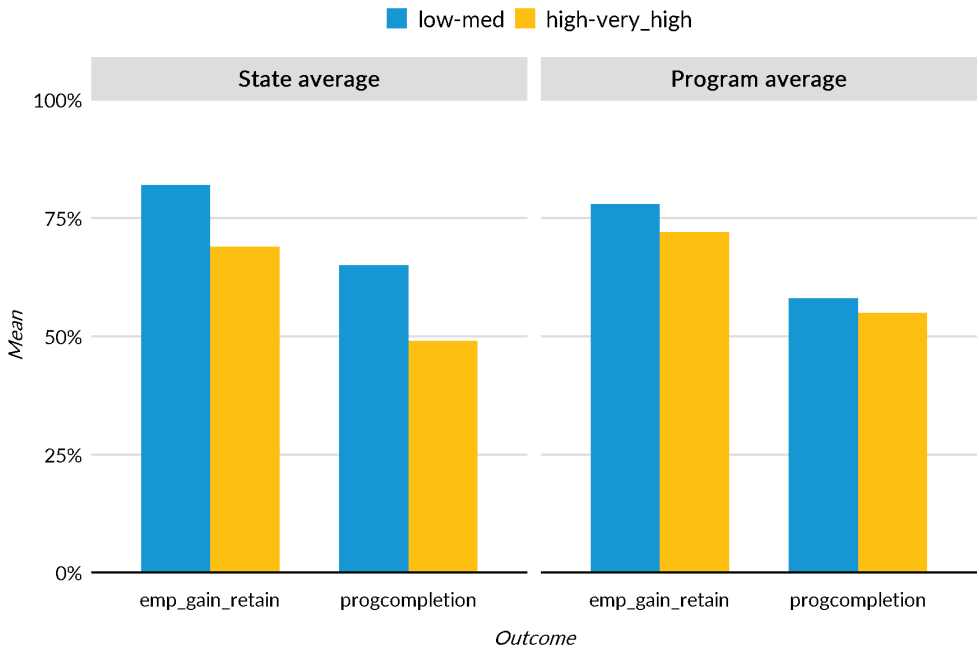


FIGURE A.2

Average LSI Change by Risk Level (FY 14-18)

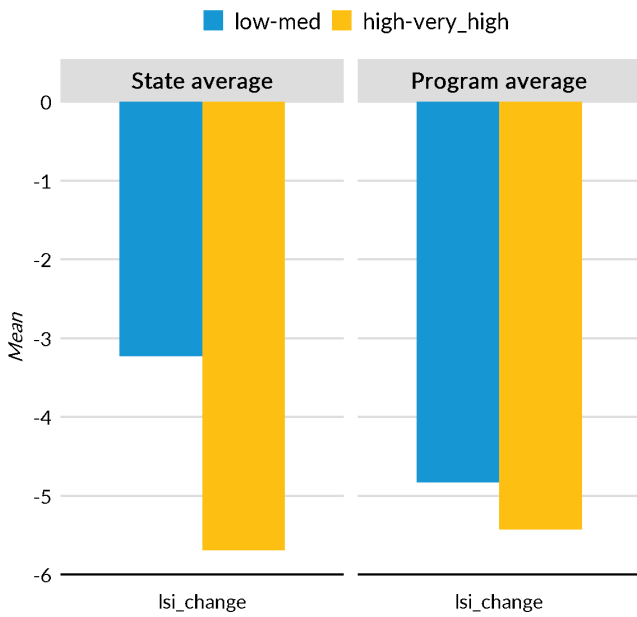
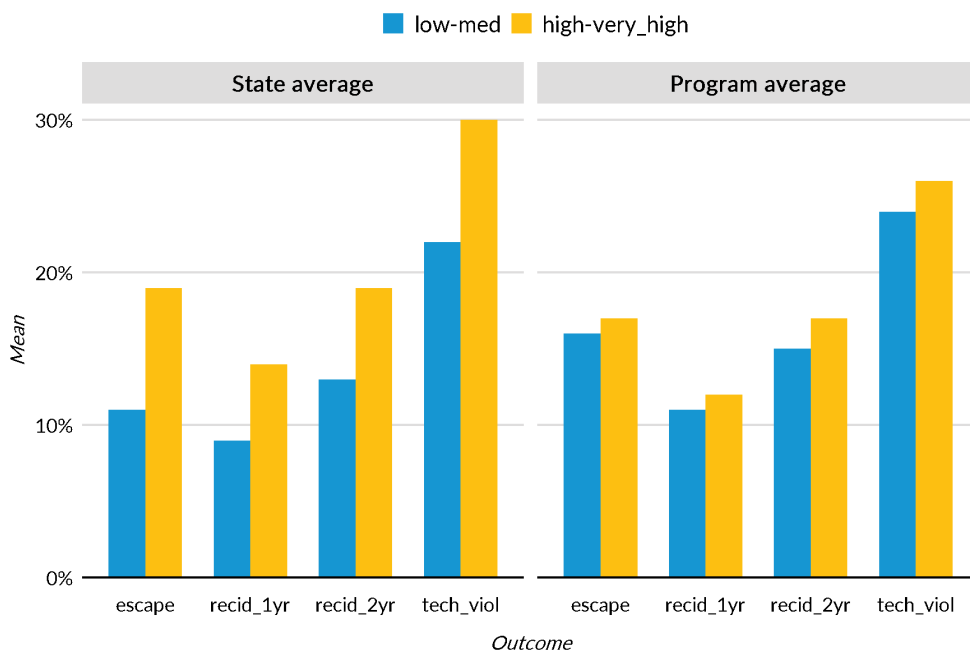


FIGURE A.3

Average Negative Outcomes by Risk Level (FY 14-18)



Note: Recidivism only uses FY 2014-17.

Based on the examination of outcomes at the state level, we see that low- and medium-risk clients have significantly different averages for each outcome from high- and very high-risk clients. However, when programs are categorized as low/medium and high/very high, there are no significant differences between their averages for each outcome. In other words, stratifying clients by risk level and calculating their average outcomes shows larger differences than when stratifying programs by risk level. No program only has clients who are high/very high risk or low/medium risk, so it is not surprising that there would be smaller differences between programs when stratified by risk.

Using the risk-adjusted outcomes above, we calculated performance targets that account for variation in risk. For the state-level method, we set the target for low/medium risk as the mean for all clients in the state. We did this because as previously mentioned, no program only has clients at low/medium risk. For the high-risk/very high-risk target, we set the target as the mean for all high- and very high-risk clients in the state. For the program-level method, the low/medium target is the mean for low- and medium-risk programs and the high/very high target is the mean for the high- and very high-risk programs.

For positive outcomes, the target is for the program to have a score that is equal to or greater than the mean. For negative outcomes, the target is to have a score that is equal to or less than the mean. For

example, a high-risk program would need to have a successful completion rate of 49 percent or higher to meet the target using the state-level method. For the program-level method, the target would be 55 percent or higher for high-risk programs.

TABLE A.8

Risk-Adjusted Outcome Performance Targets

State-level method

Measure	Low/medium	High/very high
Felony conviction 1yr from start	12%	14%
Felony conviction 2yr from start	17%	19%
Employment gain/retain	75%	69%
Escape rate	16%	19%
Change in LSI score	-4.6	-5.7
Successful completion rate	55%	49%
Technical violation rate	27%	30%

TABLE A.9

Risk-Adjusted Outcome Performance Targets

Program-level method

Measure	Low/medium	High/very high
Felony conviction 1yr from start	11%	12%
Felony conviction 2yr from start	15%	17%
Employment gain/retain	78%	72%
Escape rate	16%	17%
Change in LSI score	-4.8	-5.4
Successful completion rate	57%	55%
Technical violation rate	24%	26%

Program Risk Level Categorization

To compare program performance against the risk-adjusted outcome targets, the programs needed to be categorized by risk level. The risk profile of the clients in each program varies and no program has a homogenous population, making it difficult to categorize the programs by risk. We propose two methods for categorizing the programs by risk. In the first method, programs would be categorized based on the proportion of clients that are high and very high risk. If 50 percent or more of clients are high or very high risk, the program would be categorized as high/very high. Otherwise, the program

would be categorized as low/medium because 50 percent or more of its clients are low or medium risk. In the second method, programs are categorized based on the average risk score. If a program's average risk score is 28 or lower, the program is low/medium risk. If the average risk score is 29 or higher, the program is high/very high risk. Clients that are missing an initial LSI score are excluded from these calculations. Table A.10 shows how many programs would fall into each risk group in FY 2019. With either method, the majority of programs fall into the high/very high group. The methods are extremely similar, with the first method allowing one additional program to fall into the low/medium group.

TABLE A.10
Risk Categorizations of Programs (FY 2019)

	2. Low / Medium	2. High / Very High	Total
1. Low / Medium	7	2	9
1. High / Very High	1	23	24
Total	8	25	33

Program Performance against Targets

With the performance targets set and the programs categorized by risk, the final step is to determine whether the programs are meeting the risk-adjusted outcome targets. We compare program performance in FY 2019 to the performance targets set using the historical FY 2014–18 data. For the recidivism outcomes, we compare program performance in FY 2018 to the performance targets set using the historical FY 2014–17 data. We show the proportion of programs that meet each outcome target by risk group. We do not list out individual programs.

We have outlined two sets of performance targets and two sets of risk categorizations for the programs. We examined program performance against the state-level targets using the first risk categorization method (figures A.4 and A.5). We also examined program performance against program-level targets using the second risk categorization method (figures A.6 and A.7). The first method shows that 67 percent of low- and medium-risk programs and 83 percent of high- and very high-risk programs meet the successful completion target (figure A.4). For felony reconviction within two years, 25 percent of low- and medium-risk programs meet the target, whereas 68 percent of high- and very high-risk programs meet the target (figure A.5). For the program-level adjustment method, 38 percent of low- and medium-risk programs and 68 percent of the high- and very high-risk programs meet the successful completion target (figure A.6). For felony reconviction within two years of program start, 25 percent of low- and medium-risk and 50 percent of high- and very high-risk programs meet the target (figure A.7).

FIGURE A.4

Proportion of Programs Meeting Targets

State-level targets, first risk categorization method

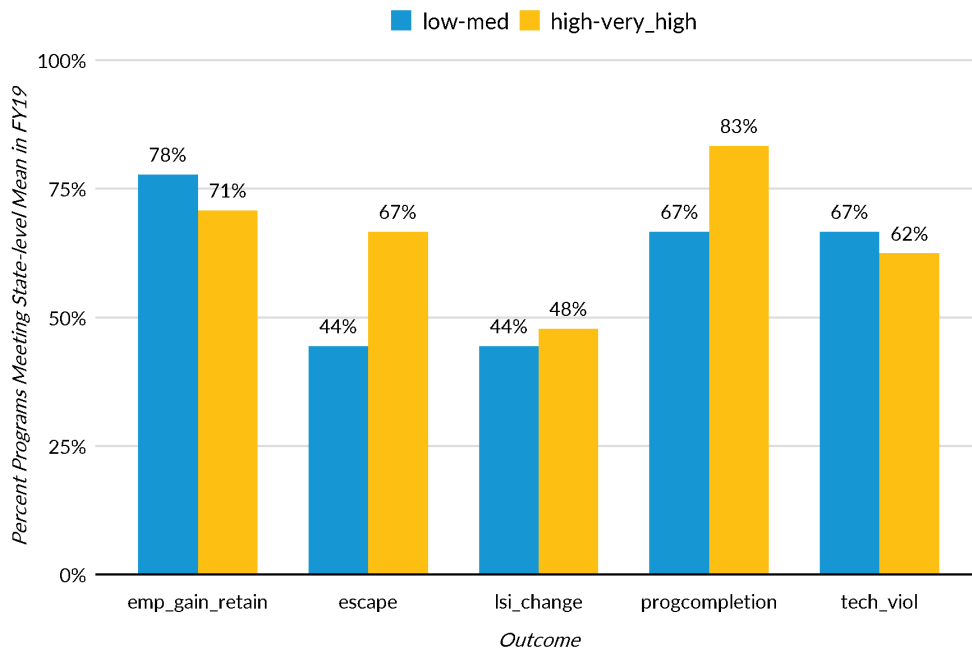


FIGURE A.5

Proportion of Programs Meeting Recidivism Targets

State-level targets, first risk categorization method

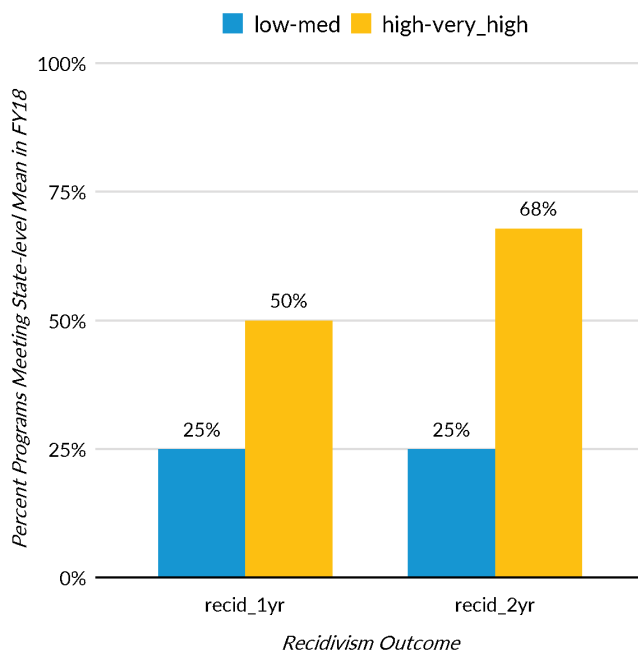


FIGURE A.6

Proportion of Programs Meeting Targets

Program-level targets, second risk categorization method

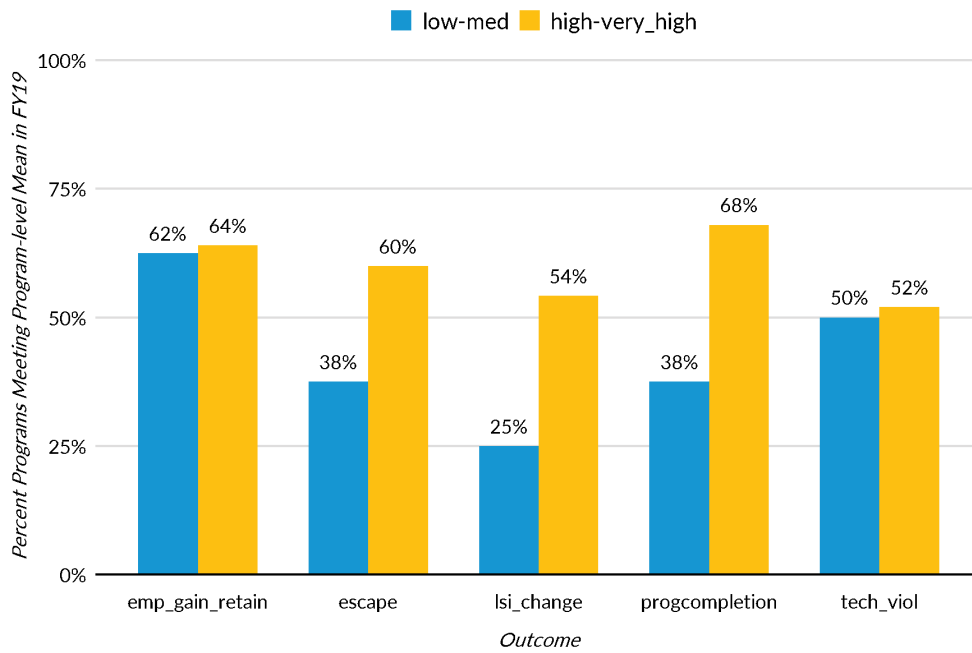
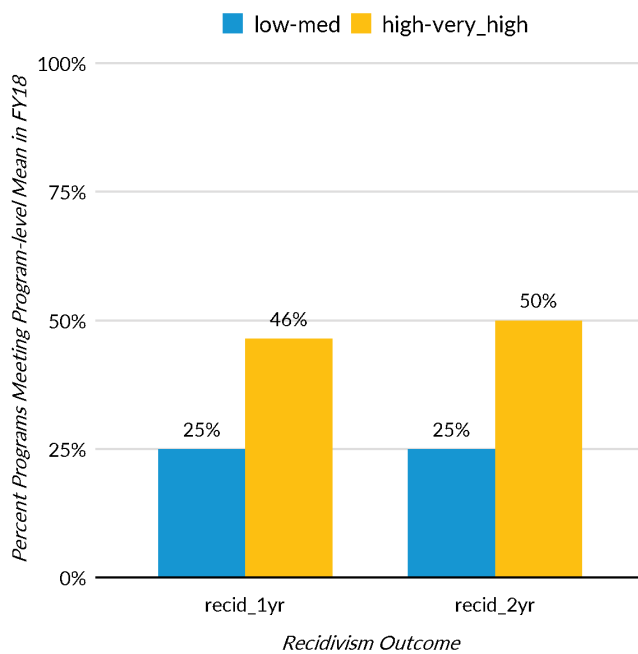


FIGURE A.7

Proportion of Programs Meeting Recidivism Targets

Program-level targets, second risk categorization method



Validity and Reliability

After we generated the outcome variables, we examined the validity and reliability of the outcomes we had created as well as the reliability of the PACE and the Core assessments. Through this analysis, we hoped to answer two questions: Are the outcomes/tools measuring what they intend to measure? And do the outcomes/factors/domains load well together? We examined the correlation between outcome measures at the program level and Cronbach's alpha to answer these questions.

Correlation

We generated a correlation matrix to measure the convergent validity of the outcome measures, or how related the outcomes are to each other. Correlation coefficient values range from -1 to 1, where -1 indicates there is a perfect negative correlation between two variables, 1 indicates perfect positive correlation, and 0 indicates that there is no linear relationship between the variables. Generally, a coefficient of 0.4 or above indicates a strong association between two variables. Table A.11 shows a correlation matrix of all of the generated outcomes. The correlation matrix indicates that a number of outcomes appear to be correlated. For example, with a coefficient of 0.63, successful completion and employment gain/retention are highly correlated outcomes. The PACE and Core results are not included in this analysis because there are not enough observations of each to conduct a meaningful correlation test.

TABLE A.11
Correlation between Outcome Measures (FY 2014–18, n=166)

Outcome	1	2	3	4	5	6	7
1. Successful completion rate	1						
2. Employment gain/retention	0.63***	1					
3. Technical violation rate	-0.70***	-0.40***	1				
4. Escape rate	-0.56***	-0.40***	-0.02	1			
5. LSI change	-0.03	-0.09	0.07	-0.11	1		
6. Felony conviction 1yr from start	-0.53***	-0.27***	0.17*	0.60***	0.05	1	
7. Felony conviction 2yr from start	-0.44***	-0.21***	0.13	0.51***	0.03	0.93***	1

Notes: **p < 0.05; ***p < 0.01.

Cronbach's Alpha

While correlation tells us the relationship between two outcomes, examining Cronbach's alpha provides us with a measure of internal consistency, or how closely related a set of items are as a group.

Cronbach's alpha tests the intercorrelations among items in a test or a survey, indicating how well a set

of items measures a single construct or concept. In other words, it helps us to understand how well certain factors load together. Cronbach's alpha coefficients range from -1 to 1. A Cronbach's alpha of 0.7 or higher is considered acceptable (Taber 2018). A negative alpha indicates that one of the items in the factor may have an opposite scale from the others. Table A.12 includes the alpha values and bounds for PACE, Core, and the generated positive and negative outcomes.

TABLE A.12
Cronbach's Alpha Values

Factor	Items	Alpha	Lower Bound	Upper Bound
PACE (n = 29)				
All Factors	7	0.72	0.55	0.81
F1 - Risk/Need	2	0.31	-0.36	0.70
F2 - Motivation	3	0.68	0.49	0.80
F3 - Target Intervention	4	0.64	0.33	0.78
F4 - Skill Train	3	0.48	0.00	0.72
F5 - Positive Reinforcement	3	-0.12	-1.17	0.44
F6 - Violation Response	3	0.56	0.31	0.71
F7 - Ongoing Support	3	0.13	-0.98	0.53
Core Security (n = 24)				
All Factors	14	0.53	-0.67	0.78
Outcomes FY 2014-18 (n=166)				
Positive outcomes	2	0.78	0.68	0.48
Negative outcomes	4	0.42	0.22	0.57

Notes: A Cronbach's alpha of 0.7 and above indicates acceptable internal consistency reliability. The lower and upper bounds refer to the lower and upper bounds of a 95 percent confidence interval. LSI change is not included in the outcomes as it has a different scale than the other outcomes.

Calculations

In this section, we describe how to calculate the performance targets going forward. For this baseline assessment, we used historical data: FY 2014-18 to calculate the targets and FY 2019 to compare program performance. For recidivism, FY 2014-17 were used to develop the target and FY 2018 was used to compare program performance. Going forward, a similar lag will be necessary. For example, we recommend using FY 2016-20 to create the target and FY 2021 to compare performance for payments in FY 2022. For recidivism, FY 2015-19 would be used for the targets and FY 2020 for comparing performance. The general pattern is that five years of data should be used for calculating the targets so that the targets adjust as the community corrections population changes over time. For the administrative outcomes, there will be a one-year lag and for the recidivism outcomes, there will be a two-year lag.

All code for calculating the targets and comparing program performance is available publicly on GitHub.⁷ In future years, this code could be updated with the desired year ranges. Because these calculations are based on administrative data, accurate and consistent data entry by programs will be extremely important. We also have provided the code for conducting the Cronbach's alpha and correlation analyses. As more PACE and Core assessments are completed, these analyses should be replicated to assess their reliability and validity.

Notes

- ¹ Rarely are contract payments based solely on the achievement of outcomes because doing so would pose a major risk to providers because outcomes take time to be realized and are not always under the control of providers. Instead, a large portion of payment is assigned to activities or quality measures within direct control of providers.
- ² “Performance-Based Contracting in Tennessee’s Foster Care System,” Center for State Child Welfare Data, accessed October 8, 2020, https://fcd.chapinhall.org/wp-content/uploads/2018/02/Performance-Based-Contracting-in-Tennessee_2.2018.pdf.
- ³ Pennsylvania Department of Corrections, “Performance-Based Contracts Continue to Positively Affect Recidivism,” *PR Newswire*, August 25, 2015, <https://www.prnewswire.com/news-releases/performance-based-contracts-continue-to-positively-affect-recidivism-300132906.html>.
- ⁴ Glenn A. Tapia, letter to boards and community corrections stakeholders, July 10, 2017, retrieved from <https://cdpsdocs.state.co.us/occ/Standards/LettertoBoardsandCCStakeholders-2017StandardsandPACE-AsPublished.pdf>
- ⁵ In addition, a small subset of individuals were coded with a termination reason of “0,” which does not have a corresponding codebook value. It was determined these individuals likely fit under the “Other” category and thus were also excluded.
- ⁶ The Colorado Department of Public Safety considers an LSI score of 29 or higher high risk. For more information, see: https://cdpsdocs.state.co.us/ccjj/Meetings/2008/2008-05-09_LSI.pdf.
- ⁷ Code for calculating the outcome measure targets, evaluating program performance, and assessing the Core and PACE is available at <https://github.com/UrbanInstitute/co-comm-corrections-pbc>. All code is in the R programming language.

References

- Governor's Community Corrections Advisory Council. 2015. *Performance Based Contracting (PBC) for Colorado Community Corrections: General Attributes of a PBC Plan*. Denver: Colorado Division of Criminal Justice, Office of Community Corrections.
- Hawkins, Rayanne, Brian Bieretz, and Madeline Brown. 2019. *Incentivizing Results: Contracting for Outcomes in Social Service Delivery*. Washington, DC: Urban Institute.
- Martin, L, and Jon Desenberg. 2009. *Moving to Performance-Based Contracting (PDC): A Report Prepared for the Michigan Department of Corrections*. Washington, DC: The Performance Institute.
- Taber, Keith S. 2018. "The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education." *Research in Science Education* 48 (6): 1,273–96.

About the Authors

Rayanne Hawkins is a policy program manager in the Research to Action Lab at the Urban Institute where she researches performance-based contracting, pay for success, and apprenticeship programs. Formerly affiliated with the Urban Institute's Justice Policy Center, she has studied best practices for criminal justice interventions involving substance use, behavioral health, and housing. She received her MBA from Georgetown University.

Ashlin Oglesby-Neal is a research associate in the Justice Policy Center where her research includes developing and validating risk assessment tools, and evaluating criminal justice policies and programs. Her work includes multisite data collection and management, collaboration with criminal justice practitioners, quantitative data analysis, and data visualization. She received a MS in Criminology from the University of Pennsylvania.

Libby Doyle is a research assistant at the Justice Policy Center. Her research focuses on evaluations of community-based programs, reentry, and public defense. She received a BA in Political Science and Public Policy from the University of North Carolina at Chapel Hill.

Isabella Remor is a policy assistant in the Research to Action Lab, where she works on a variety of projects in the fields of inclusive growth, housing, and upward mobility. Before joining Urban, she worked at a public relations firm to promote clean energy strategies and advocate for bail reforms in New York and Mississippi. She received her BA in political science from New York University.

STATEMENT OF INDEPENDENCE

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.



500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org